Infrared spectroscopy as a clinical diagnostic method for detection of disease states: developments and applications in kidney diseases and cancer diagnoses

by

Katherine Victoria Oliver

Submitted for the degree of Doctor of Philosophy

in

Structural, Computational and Chemical Biology

University College London

Thesis committee:

Chair: Dr. Katherine Thompson
Principal investigator: Professor Peter Rich
Co-investigator: Professor Robert Unwin

2015

Signed declaration

I, Katherine Victoria Oliver, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

_____

Date

Abstract

ATR-FTIR spectroscopy has been used to develop rapid and accurate methods for urine analyses and cancer diagnoses.

Complications associated with analyses of dried samples by FTIR spectroscopy were found to arise because of the existence of partial hydration states of some substances. Recognition and accommodation of such states is necessary to ensure precise deconvolution of spectra. Dramatic hydration effects were observed in the IR spectrum of urea. Four forms of urea were identified in IR spectra, corresponding to solution (1468 cm^{-1}) and dry (1464 cm^{-1}) states and two intermediate (1454 cm^{-1} and 1443 cm^{-1}) forms that arise from specific urea-water and/or urea-urea interactions. Less pronounced hydration effects were also identified in the spectra of creatinine and other urinary compounds.

ATR-FTIR spectroscopy was then applied to diagnosis of the kidney stone disease, cystinuria. A simple method for detecting and quantitating urinary insoluble cystine was developed based on the integrated area of the cystine 1296 cm^{-1} band. Clinically relevant levels could be detected in ATR-FTIR spectra of dried insoluble fractions of urine samples without any requirement for chemical manipulation.

The use of ATR-FTIR spectroscopy as a point-of-care diagnostic tool for kidney diseases was investigated by installation of a portable benchtop spectrometer in a hospital clinic. Nurses recorded spectra of 295 unprocessed urine samples and the data were correlated with clinical diagnoses. Initial results suggested that differences in the integrated areas of key spectral features could discriminate between healthy patients and those with chronic kidney disease (CKD), and indicate the extent of kidney damage.

ATR-FTIR spectra of 80 pancreatic tissue samples were also recorded and analysed for spectral patterns associated with cancer using PLS logistic regression. Tumour tissue spectra could be separated from normal tissue spectra with 90 % sensitivity. Tumour samples could be further separated into cancer stages with accuracies of 64-84 %.

Contents

List of acronyms **9**

Notation and symbols **11**

List of Figures **16**

List of Tables **20**

1 Introduction **22**

1.1 Introduction to infrared spectroscopy 23

1.1.1 Fourier transform and the Michelson interferometer 24

1.1.2 Attenuated total reflection (ATR) 24

1.1.3 Pros and cons of ATR and transmission 27

1.1.4 Line shapes in infrared spectra 28

1.1.5 FTIR microspectroscopy 28

1.2 Raman spectroscopy 30

1.3 Data analysis methods 31

1.3.1 Feature extraction 31

1.3.2 Unsupervised and supervised classification methods 32

1.4 FTIR spectroscopy applied to medical diagnostics 33

1.4.1 Biofluids 33

1.4.1.1 IR analyses of blood and blood products 34

1.4.1.2 IR analyses of urine 36

1.4.1.3 Enhancing sensitivity of IR measurements of biofluids 37

1.4.1.4 IR analyses of kidney stones 38

1.4.2 Tissue analyses and cancer diagnostics 39

1.4.2.1 Cervical and prostate cancers . . . 39
1.4.2.2 Lymph nodes . . . 40
1.4.2.3 Barrett's oesophagus . . . 40
1.4.2.4 Lung cancer . . . 40
1.4.2.5 Breast and ovarian cancers . . . 41
1.4.2.6 Brain cancers . . . 42
1.5 Aims . . . 42

2 Materials and methods 46

2.1 Materials . . . 47
2.1.1 Chemicals . . . 47
2.1.2 Clinical samples . . . 47
2.2 Spectroscopic methods . . . 48
2.2.1 ATR-FTIR and transmission FTIR spectroscopy . . . 48
2.2.2 FTIR spectroscopy using a benchtop FTIR spectrometer in the clinic . . . 49
2.2.3 Transmission FTIR spectroscopy of volatiles . . . 50
2.3 Sample handling . . . 50
2.3.1 ATR-FTIR measurements of liquid samples . . . 50
2.3.2 Drying samples and dealing with insoluble material for ATR-FTIR spectroscopy . . . 51
2.3.3 Separating insoluble material . . . 52
2.3.4 ATR-FTIR measurements of solids . . . 52
2.3.5 Monitoring spectral changes associated with hydration states . . . 52
2.3.5.1 Controlled rehydration of urea . . . 52
2.3.5.2 Serial dehydration of urinary components and urine . . . 53
2.3.6 ATR-FTIR measurements of tissues . . . 53
2.3.7 Vapour phase measurements of volatile samples in transmission mode . . . 54
2.4 Data processing and analyses . . . 54
2.4.1 Data preprocessing . . . 54
2.4.2 Ramping of ATR-FTIR spectra . . . 57
2.4.3 Curve fitting . . . 57

2.4.4 Integration analysis . . . 58
2.4.5 Bland-Altman analysis . . . 59
2.4.6 Partial least squares (PLS) logistic regression . . . 59
2.4.7 Calculation of sensitivities and specificities . . . 59
2.5 Electronic structure modelling of urea-water and urea-urea structures . . . 60
2.6 Urine analyses . . . 60
2.6.1 Generation of urea and creatinine calibration curves . . . 60
2.6.2 Generation of a cystine calibration curve . . . 60
2.6.3 Quantitation of creatinine by the Jaffe reaction . . . 61
2.6.4 Standard clinical quantitation of cystine . . . 61
2.6.5 Vapour phase measurements of trimethylamine (TMA) . . . 61
2.6.6 Solution phase measurements of TMA and TMA N-oxide . . . 62

3 Model compounds 63
3.1 Introduction . . . 64
3.2 Urine-related compounds . . . 64
3.3 Tissue-related compounds . . . 73
3.4 Discussion and conclusions . . . 83

4 Hydration effects on infrared spectra 86
4.1 Introduction . . . 87
4.2 Transmission IR spectra of urea . . . 89
4.3 Kinetic resolution of multiple urea hydration states . . . 90
4.4 Stepwise rehydration of urea . . . 92
4.5 Component fitting to urea spectra . . . 93
4.6 Hydration states of creatinine . . . 96
4.7 Hydrated forms of urea in dried urine samples . . . 99
4.8 Factors in urine affecting observed behaviour . . . 104
4.9 Electronic structure modelling and physical basis of intermediates . . . 105
4.10 Hydration effects of other compounds in urine . . . 106
4.10.1 Ammonium chloride . . . 106
4.10.2 Sulphate . . . 107
4.10.3 Phosphate . . . 108

4.10.4 Glucose . . . 109
4.10.5 Cystine . . . 110
4.11 Discussion and conclusions . . . 111

5 Analyses of urine FTIR spectra for disease diagnosis. 114
5.1 Case study: cystinuria . . . 115
5.1.1 Introduction . . . 115
5.1.2 Quantitation of urea and creatinine in urine by ATR-FTIR spectroscopy . . . 117
5.1.3 Comparison of ATR-FTIR and Jaffe reaction methods for creatinine quantitation . . . 123
5.1.4 Development of measurement of insoluble compounds in urine . 124
5.1.5 Analysis of cystinuric urine samples . . . 126
5.1.6 Comparison of FTIR and clinical methods of cystine quantitation 129
5.2 Application of methods to other urinary disease biomarkers. Case study: APRT-deficiency . . . 130
5.2.1 Introduction to APRT-deficiency . . . 131
5.2.2 Analysis of APRTd patient stone and urine samples . . . 131
5.2.3 Analysis of a transmission spectrum of urine from a suspected APRTd patient . . . 138
5.3 Case study: trimethylaminuria . . . 139
5.3.1 Introduction to trimethylaminuria . . . 139
5.3.2 Vapour and solution phase spectra of trimethylamine . . . 140
5.4 Discussion and conclusions . . . 144

6 ATR-FTIR spectroscopy as a point-of-care diagnostic tool for detecting disease biomarkers in fresh urine samples 151
6.1 Introduction . . . 152
6.2 Development of protocols for measuring urine samples on the benchtop spectrometer . . . 153
6.3 Final dataset after quality control . . . 154
6.4 Identification of spectral variation and comparison to model compound spectra . . . 156

6.5 Comparison of all renal-related disease urine spectra with healthy urine spectra . . . 157
6.6 Spectral identification of urine from patients with abnormal glomerular filtration rate (GFR) . . . 161
6.6.1 Patients with estimated GFR below 60 mL/min (CKD stage 3+) . 161
6.6.2 Patients with eGFR below 30 mL/min (CKD stage 4+) . . . 165
6.7 Discussion and conclusions . . . 168

7 Detection of pancreatic cancer by ATR-FTIR spectroscopy 171
7.1 Introduction . . . 172
7.2 Summary of ATR-FTIR spectra collected of pancreatic resections . . . 174
7.3 Integration analysis of healthy and cancerous pancreas biopsy spectra . . 176
7.4 PLS-logistic regression model: staging of tumour tissue . . . 178
7.4.1 Performance of model 1: separating all normal spectra from all tumour spectra . . . 180
7.4.2 Performance of model 2: separating stage 3 tumour spectra from stages 0-2 tumour spectra . . . 183
7.4.3 Performance of model 3.: separating stage 2 tumour spectra from stages 0 & 1 tumour spectra . . . 185
7.5 Correlation of pancreas biopsy spectra with model compound spectra . . 188
7.6 Discussion and conclusions . . . 191

8 Concluding remarks and future prospects 194

Bibliography 199

A Benchtop spectrometer instructions 218

List of acronyms

A	absorbance
ADC	adenocarcinoma
AKI	acute kidney injury
ANN	artificial neural networks
APRTd	adenine phosphoribosyltransferase deficiency
ATP	adenosine triphosphate
ATR	attenuated total reflection
BSA	bovine serum albumin
CA	carbohydrate antigen
CKD	chronic kidney disease
CKD-EPI	chronic kidney disease epidemiology collaboration
CT	computed tomography
DA	discrimination analysis
DHA	dihydroxyadenine
DNA	deoxyribonucleic acid
DTGS	deuterated triglycine sulfate
eGFR	estimated glomerular filtration rate
ESRF	end stage renal failure
FTIR	Fourier transform infrared
FWHM	Full Width Half Maximum
GFR	glomerular filtration rate
HCA	hierarchical cluster analysis
IR	infrared
LDA	linear discriminant analysis

LFDI	laminar fluid diffusion interface
LV	latent variable
M1	fixed mirror
M2	moving mirror
MCT	mercury cadmium telluride
MDRD	modification of diet in renal disease
MLR	multinomial logistic regression
mM	millimolar
MRI	magnetic resonance imaging
nm	nanometre
P	patient
PC	principal component
PCA	principal components analysis
PLS	partial least squares
pNET	pancreatic enuroendocrine tumour
RFH	Royal Free Hospital
SNR	signal to noise ratio
SNV	standard normal variate
SVM	support vector machine
TMA	trimethylamine
TNM	Tumour Nodes Metastasis
UCL	University College London
UTI	urinary tract infections
WDN	wavelet denoising

Notation and symbols

δ	bend
ϵ	extinction coefficient
θ	angle of incidence
λ	wavelength
μ	micro
ν_{as}	antisymmetric stretch
ρ	rock
τ_a	amplitude correlation time
τ_c	coherence lifetime
cm^{-1}	wavenumber
I	intensity of incident light
I_0	intensity of light from sample at a specific frequency wavelength
KBr	potassium bromide
KCl	potassium chloride
KNO_3	potassium nitrate
m_u	profile shape factor
n_1	refractive index of ATR prism material
n_2	refractive index of sample
NaCl	sodium chloride
$NaNO_3$	sodium nitrate
NH_4Cl	ammonium chloride
NH_4SO_4	ammonium sulphate
x_c	peak centre
y_0	baseline centre

For Tom

Acknowledgements

Firstly, I would like to express my sincerest gratitude to my supervisor, Professor Peter Rich, for his support, guidance and patience over the course of my Ph.D studies, and for sharing his extensive knowledge through insightful discussions and critical evaluation of my work.

My thanks also go to my secondary supervisor, Professor Robert Unwin, for his expert clinical input to the urine analyses projects, and for the advice and probing questions he has offered during my committee meetings. Thanks to him also for providing urine samples for the cystinuria and DHA projects, and for facilitating the installation of the benchtop spectrometer at the Royal Free Hospital (RFH).

I would like to thank my thesis committee chair, Dr. Katherine Thompson, for her encouraging feedback in each of my committee meetings.

Thanks to present and former group members for sharing their expertise, offering their advice and for providing an enjoyable working atmosphere. To Dr. Amandine Maréchal for her valuable contributions to the urea hydration and cystinuria projects, and general critical scientific input during lab meetings. To Talha Arooz, for exceptional research assistance, and for always going above and beyond to lend a hand. Also, for his unjudgemental encouragement and positive attitude. To former group member, Dr. Raksha Dodia, for sharing her knowledgeable opinion whenever a problem arose and for the laughter shared in and out of the lab.

Special thanks to fellow group member, Liberty Foreman, for her technical computing assistance, for sharing her ideas and providing feedback on all aspects of my work. Most of all, for her friendship and ambition that have steered me through to the end of this Ph.D and on to the next exciting challenge.

Thanks to former technician, Santi Garcia, for his ability to build anything and for his comedy.

I am grateful to Dr. Shabbir Moochhala for assisting in the installation of the benchtop spectrometer at the RFH and for his clinical guidance in the urine analyses projects. To Faith Matjiu and Cameron Davey for their unfailing enthusiasm and dedication in recording the urine spectra and collating clinical data at the Royal Free Hospital, in spite of a heavy workload and shortage of staff.

I acknowledge Dr. Karen Yu, Dr. Richard Fish, Emma Miler, Sarah Mappleback and Dr. Vidar Edvardsson for provision of urine samples. Also, Dr. Annalisa Vilasi for providing preliminary data for the cystinuria project.

I express my gratitude to Professor Graça Soveral and Dr. Jorge Paulino for

provision of the pancreatic tissue samples and for their clinical input to the pancreatic cancer project.

I acknowledge that the research would not have been possible without the financial support from the Medical Research Council.

Finally, I offer my deepest gratitude to my friends and family, for continual love and support. To my parents for their infallible advice and unconditional support. To my sisters, Ali and Jenny, and brothers-in-law, Justin and Steve, for their friendship and shared laughter. And to my partner, Tom, for his love, patience and humour through the highs and lows of the PhD.

Publications

Oliver KV, Vilasi A, Maréchal A, Moochhala SH, Unwin RJ and Rich PR, 'Infrared vibrational spectroscopy as a diagnostic tool for cystinuria'. In preparation.

Oliver KV, Maréchal A and Rich PR (2015), 'Effects of hydration state on the mid-infrared spectra of urea and creatinine in relation to urine analyses'. *Applied Spectroscopy*. Accepted.

Oliver KV, Matjiu F, Davey C, Moochhala SH, Unwin RJ and Rich PR (2015), 'Attenuated total reflection Fourier transform infrared (ATR-FTIR) spectroscopy as a bedside diagnostic tool for detecting renal disease biomarkers in fresh urine samples'. *Proceedings of SPIE 9332, Optical Diagnostics and Sensing XV: Toward Point-of-Care Diagnostics*,

List of Figures

1.1 Schematic of the Michelson interferometer. 25
1.2 Interferogram and power spectrum. 25
1.3 Schematic of the ATR prism 26
1.4 Lorentzian and Gaussian line shapes. 29

2.1 Bruker IFS/66S spectrometer. 49
2.2 Bruker Alpha FTIR spectrometer. 50
2.3 Drying a sample on to the ATR prism with a gentle stream of dry nitrogen gas. 51
2.4 Schematic of the humidification chamber device. 52
2.5 Vapour chamber device for transmission measurements of volatiles. . . . 54
2.6 Water vapour and liquid water spectra. 56

3.1 Model spectra of urea. 66
3.2 Model spectra of creatinine. 67
3.3 Model spectra of ammonium ions. 68
3.4 Model spectra of potassium phosphate at pH 6. 69
3.5 Model spectra of potassium sulphate at pH 7. 70
3.6 Model spectra of glucose. 71
3.7 Model spectra of selected stone-forming compounds. 73
3.8 Model spectrum of solid bovine serum albumin. 74
3.9 Model spectra of collagen. 75
3.10 Model spectra of creatine. 76
3.11 Model spectra of phosphatidylcholine. 77
3.12 Model spectra of cholesterol. 78
3.13 Model spectra of glycogen. 79

3.14 Model spectra of RNA and DNA. 80
3.15 Model spectra of mucin. 81
3.16 Model spectra of hydroxyproline. 82
3.17 Model spectra of selected poly-amino acids. 83

4.1 Variation in the v_{as}(CN) band of urea in wet and dried pure urea and urine spectra. 89
4.2 Transmission spectra of urea. 90
4.3 Serial spectra of a rehydrated urea sample during drying. 91
4.4 Solution and crystalline spectra of urea. 92
4.5 Spectra of the v_{as}(CN) band of urea between 1510 and 1410 cm^{-1} at fixed relative humidities. 93
4.6 Component fitting to the v_{as}(CN) band of urea during the transition from rehydrated to dry states. 95
4.7 Component fitting to the v_{as}(CN) band of urea in solution. 96
4.8 Serial spectra of a rehydrated urea sample during drying. 97
4.9 Component fitting to the 1520 and 1470 cm^{-1} spectral region of creatinine during the transition from rehydrated to dry states. 98
4.10 Serial spectra of rehydrated urine samples during drying. 100
4.11 Serial spectra of the v_{as}(CN) band of urea in four additional urine samples during drying. 101
4.12 Fitting of urea and creatinine components to the 1510–1410 cm^{-1} region of spectra of human urine. 102
4.13 ATR-FTIR spectrum of a typical urine sample. 103
4.14 Behaviour of the v_{as}(CN) urea band in dried 'simulated urine'. 104
4.15 Serial spectra of a rehydrated ammonium chloride sample during drying. 107
4.16 Serial spectra of a rehydrated potassium sulphate sample during drying. 108
4.17 Serial spectra of a rehydrated potassium phosphate sample during drying. 109
4.18 Serial spectra of a rehydrated glucose sample during drying. 110
4.19 Serial spectra of a rehydrated cystine sample during drying. 111
4.20 Modelled urea-urea and urea-water structures. 112

5.1 ATR-FTIR spectra of urea and creatinine and deconvolution of mixtures. 120

5.2 Calibration curves for urea and creatinine in water. 121
5.3 Deconvolution of urea and creatinine in whole urine spectra. 122
5.4 Urea and creatinine concentrations in urine samples. 122
5.5 Comparison of ATR-FTIR spectroscopy and Jaffe methods of creatinine quantitation in whole urine samples. 123
5.6 The effect of drying time on reproducibility of cystine spectra. 124
5.7 The effect of homogenisation and gentle pressure on the reproducibility of cystine spectra. 125
5.8 Structure of cystine. 126
5.9 Absorbance spectra of cystine and insoluble cystinuric urine. 127
5.10 Absorbance spectra of other urinary amino acids associated with cystinuria. 127
5.11 Cystine calibration curve. 128
5.12 Levels of cystine measured in cystinuric and control urine samples by ATR-FTIR spectroscopy. 129
5.13 Comparison of cystine quantitation in urine by ion exchange chromatography and by ATR-FTIR spectroscopy. 130
5.14 Structure of 2,8-dihydroxyadenine. 131
5.15 Reference absolute absorbance spectra of 2,8-DHA. 132
5.16 Second derivative of the absorbance spectrum of synthesised 2,8-DHA. . 132
5.17 IR spectra of APRTd patient urine A. 133
5.18 IR spectra of APRTd patient urine B-F. 134
5.19 IR spectra of APRTd patient urine D-F supernatants. 135
5.20 Comparison of APRTd patient urine D-F with model compounds. 136
5.21 IR spectra of urine from rat model of APRTd. 137
5.22 Transmission spectrum of urine from a patient with possible APRTd. . . 138
5.23 Structures of trimethylamine (TMA) and TMA N-oxide. 139
5.24 Vapour phase transmission spectra of TMA solutions. 140
5.25 Solution ATR spectra of TMA. 141
5.26 Solution ATR spectra of TMA N-oxide. 143
5.27 Urinary cystine to creatinine ratios. 147

6.1 Simplified buttons for evoking macros in OPUS 7. 154

6.2 Examples of urine spectra recorded in the RFH clinic. 156
6.3 Comparison of three example urine spectra with model urinary compound spectra. 157
6.4 Healthy vs. diseased urine spectra recorded in RFH clinic. 158
6.5 Healthy vs. diseased urine scatter plots. 160
6.6 CKD3+ patient urine spectra. 162
6.7 CKD3+ vs. healthy scatter plot. 164
6.8 CKD3+ vs. the rest scatter plot. 164
6.9 CKD4+ patient urine spectra. 165
6.10 CKD4+ vs. healthy scatter plot. 167
6.11 Comparison of amide II absorbance between diseased and healthy groups. 167

7.1 Individual second derivative spectra of normal and tumour pancreatic tissue. 175
7.2 Averaged second derivative spectra of normal and tumour pancreatic tissue. 175
7.3 Scatter plot comparing normal pancreatic and tumour tissue. 177
7.4 Averaged second derivative spectra of normal pancreatic tissue and tumour stages 0-3. 177
7.5 Scatter plot comparing normal pancreatic and tumour stages 0-3. 178
7.6 Proposed diagnostic pipeline for classifying FTIR spectra from pancreatic cancer patients according to tumour stage using PLS logistic regression. 179
7.7 Loadings on the first 5 LVs for PLS logistic regression model 1. 181
7.8 3D scatter plot for model 1. 182
7.9 Loadings on the first 6 LVs for PLS logistic regression model 2. 184
7.10 3D scatter plot for model 2. 185
7.11 Loadings on the first 5 LVs for PLS logistic regression model 3. 186
7.12 Scores plot for model 3. 187
7.13 Comparison of pancreatic tissue spectra with model sugar spectra. 189
7.14 Comparison of pancreatic tissue spectra with model RNA and DNA spectra. 190
7.15 Comparison of pancreatic tissue spectra with model protein spectra. . . . 191

List of Tables

2.1 Salt solutions and corresponding humidity at room temperature. 53

4.1 Parameters of components required for simulation of urea spectra between 1510 and 1410 cm^{-1} corresponding to the v_{as}(CN) band. 94
4.2 Parameters of components required for simulation of creatinine spectra between 1520 and 1470 cm^{-1} corresponding to the (v(C=N) (30) v(CN) (27) δ(NCH) (24)) band. 99
4.3 Predicted frequencies of v_{as}(CN) band of urea in combination with water or other urea molecules. 106

5.1 Final parameters for quantitative component fitting of urea and creatinine mixed solution spectra. 118
5.2 Validation of urea and creatinine component fitting. 119
5.3 Data on cystinuric patients used in study. 146
5.4 Summary of diagnostic methods developed using FTIR spectroscopy for detection of urinary disease biomarkers. 150

6.1 Ages and sexes of patients donating urine samples to study. 155
6.2 p values for comparison of 17 spectral features between healthy and diseased samples. 159
6.3 p values for comparison of 17 spectral features between healthy and CKD3+, and between CKD3+ and eGFR>60 samples. 163
6.4 p values for comparison of 17 spectral features between healthy and CKD4+, and between CKD4+ and eGFR>30 samples. 166

7.1 Summary of numbers and stages of pancreatic tumour samples received. 174

7.2 Summary of potential features selected to distinguish between pancreas and tumour samples. 176
7.3 Confusion table for model 1. 183
7.4 Sensitivities and specificities for model 1. 183
7.5 Confusion table for model 2. 185
7.6 Sensitivities and specificities for model 2. 185
7.7 Confusion table for model 3. 187
7.8 Sensitivities and specificities for model 3. 187

Chapter 1

Introduction

Vibrational spectroscopy has been used for over a century to probe molecular structures. However, it is only relatively recently that major improvements in technology have afforded the sensitivity required to study biological molecules and it is only within the last decade or so that FTIR imaging has become available [1]. Some of the most important technological developments include the interferometer, highly sensitive detectors and array detectors, powerful light sources and attenuated total reflection (ATR) technology [2,3]. Advances in computing power have also enabled rapid processing of large datasets. Here, the focus will be on the theory and application of infrared spectroscopy in medical diagnostics. For completeness, a very brief discussion on the complementary method of Raman spectroscopy will be included.

1.1 Introduction to infrared spectroscopy

Infrared spectroscopy uses light from the infrared (IR) region of the electromagnetic spectrum. Infrared light is usually subdivided roughly into three regions: near- (~14000 - 4000 cm^{-1} / 714 - 2500 nm), mid- (~4000 – 800 cm^{-1} / 2500 - 12500 nm) and far- (~800 – 100 cm^{-1} / 12500 - 100000 nm) IR. The frequency ranges of mid- and far-IR light correspond to the energy of molecular vibrational transitions and any vibrational mode that has a net change in dipole will absorb IR light energy and so will be IR active. An IR photon may be absorbed by a molecule when its energy is equal to the energy difference between two vibrational levels. Non-linear molecules have 3n − 6 and linear molecules have 3n − 5 (where n = number of atoms) vibrational modes, many of which are IR active. This means that most types of molecules can be detected with IR spectroscopy and the spectra of biological materials are rich in information but can be very complex with multiple overlapping bands that can be difficult to deconvolute. An additional problem in biological samples can be the strong absorbance bands of liquid water in the mid- and far-IR range which can mask many other signals. The contribution from water usually has to be removed, either physically by drying the sample, or computationally by subtracting a liquid water spectrum from the sample spectra. The 1800 – 800 cm^{-1} region of the mid-IR spectrum is particularly dense in IR absorbance bands and is commonly referred to as the 'fingerprint' region [2, 4].

When the frequency of infrared light matches the energy of a molecular vibrational transition, some of the photons are absorbed and the remaining light is passed to the detector. The extent of absorption is quantitatively related to the sample concentration through the well-known Beer-Lambert law (Eq. 1.1):

$$Absorbance = log_{10}\left(\frac{I_0}{I}\right) = \epsilon.c.l \tag{1.1}$$

where I_0 = reference (incident) intensity, I = sample intensity at a specific wavelength, ϵ = extinction coefficient, c = concentration and l = pathlength. The frequencies of the absorbance bands are functions of the primary and secondary structure of the molecule [5].

1.1.1 Fourier transform and the Michelson interferometer

Fourier transform infrared (FTIR) spectroscopy allows a wide range of wavelengths of light to be simultaneously passed through the sample and deconvoluted (in contrast to dispersive-type instruments that scan each wavelength individually). The major advantages of FTIR spectroscopy are that there is less variability in the background as all the data are acquired simultaneously, and the signal-to-noise ratio is inherently much higher because larger apertures can be used. To achieve this, the broadband IR light is first passed through a Michelson interferometer (Figure 1.1). This comprises two mirrors perpendicular to one another: a fixed mirror M1, and a moving mirror, M2, which moves back and forth through distance x. The plane of these two mirrors is bisected by a beam splitter that splits the incoming beam, sending 50% of the light to M1 and 50% to M2. The two beams are reflected back to the beam splitter where they recombine. Light travelling to M1 travels through the distance 2L, whilst light travelling to M2 travels through the distance 2(L+x). Therefore, when the two beams recombine they interfere because they have an optical pathlength difference of 2x. The light passes out of the interferometer, through the sample, and is focussed at the detector. Intensity of light is recorded as a function of the M2 mirror position to produce the interferogram (Figure 1.2 A). The interferogram is then Fourier transformed to generate a power spectrum (Figure 1.2 B) of intensity versus wavelength (or, more typically in IR spectroscopy, of intensity versus wavenumber). By also recording an equivalent 'background' power spectrum in the absence of sample the Beer Lambert law (Eq.1.1) can be applied to produce plots of absorbance versus wavenumber that can be used for quantitative analyses [6].

1.1.2 Attenuated total reflection (ATR)

For this project, FTIR spectroscopy was performed primarily in attenuated total reflection (ATR) mode. In ATR mode an appropriately optically-configured IR-transmitting prism is used to totally internally reflect the IR light and the sample is placed in direct

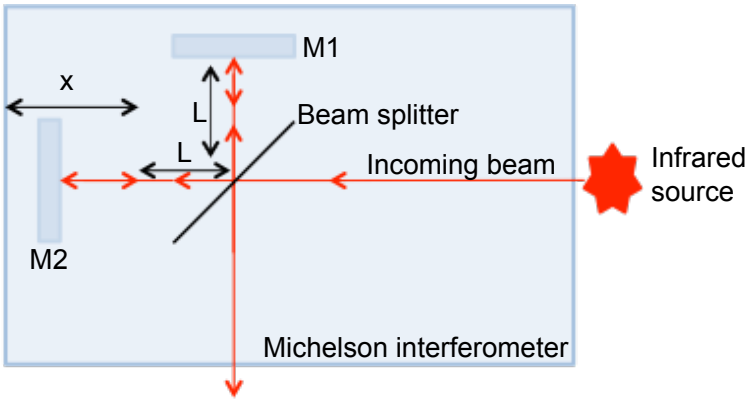


Figure 1.1: Schematic of the Michelson interferometer [1]. *M1 = fixed mirror. M2 = moving mirror, which moves through distance, x. L = distance from beam splitter to mirror.*

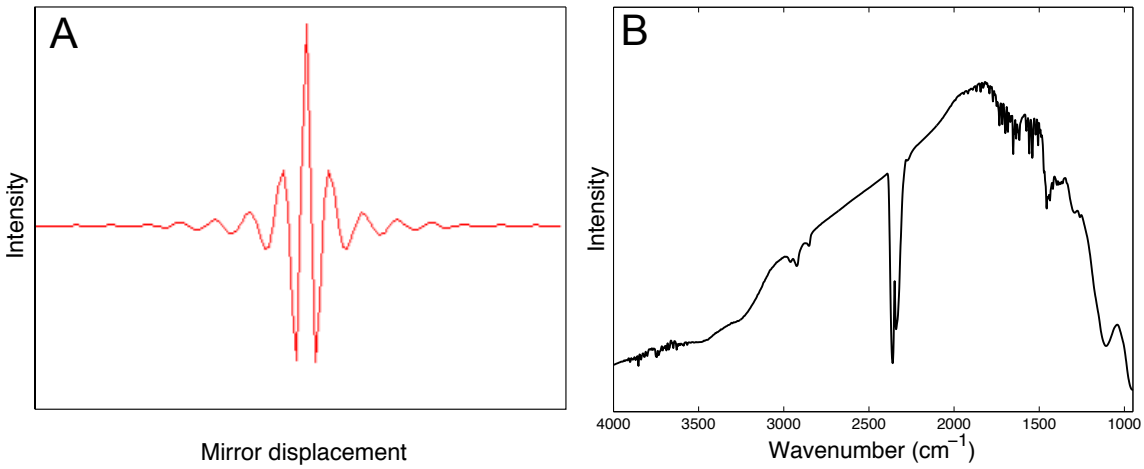


Figure 1.2: Interferogram and power spectrum. *(A) Example of an interferogram shown in the OPUS software. The intensity is greatest when the mirror displacement is 0 (M1 and M2 are equal distance from the beam splitter) as there is constructive interference when the IR beams recombine (no loss of signal). Displacement of mirror, M2, forwards or backwards results in the two IR beams becoming increasingly out of phase. Once the two IR beams are half a cycle out of phase the signal is lost due to destructive interference when they recombine. (B) Power spectrum of the blank ATR prism, generated by Fourier transforming the interferogram.*

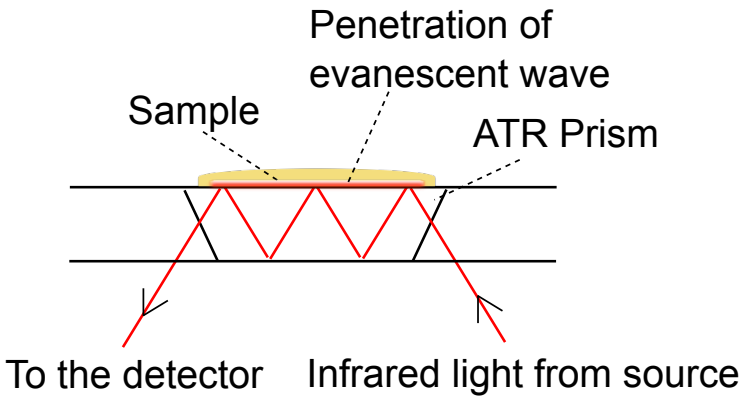


Figure 1.3: Schematic of the ATR prism. *The diagram shows how the infrared light is directed into the ATR prism where it is totally internally reflected before passing to the detector. The sample is placed directly on top of the prism as shown.*

contact with the prism surface. For total internal reflection to occur the angle of incidence of the light at the prism boundary must exceed the critical angle. The critical angle depends on the refractive indices of both the internal medium (the prism material) and the external medium (e.g. the sample). At the reflecting surface an electromagnetic disturbance extends into the external medium, called an evanescent wave. The depth of penetration of the evanescent wave into the external medium (sample) depends on the wavelength of the IR light, the angle of incidence and the refractive indices of the prism and the sample, as given by:

$$dp = \frac{\lambda_1}{2\pi(sin^2\theta - n_{21}^2)^{1/2}} \tag{1.2}$$

where dp is the depth of penetration, λ_1 is the wavelength of the incident light divided by the refractive index of the ATR prism material (λ/n_1), θ is angle of incidence of the IR beam and n_{21} is the refractive index of the external medium divided by that of the ATR prism (n_2/n_1) . Depth of penetration is generally between 0.5 and 5 µm. Some of the light is absorbed by the sample and the remaining light is passed to the detector and can be Fourier transformed to generate power and absorbance spectra as above (Figure 1.3) [3].

1.1.3 Pros and cons of ATR and transmission

ATR allows samples to be placed directly onto the prism with minimal or no sample preparation and without the need to place samples between IR-transmitting windows. This is a particular advantage with biological samples where treatment may disrupt the native biological state. For example, biopsy samples can be analysed without sectioning (as would be required for traditional transmission IR spectroscopy) and whole urine samples can be spotted directly onto the prism. The non-destructive nature of FTIR spectroscopy, and the fact that chemical manipulation or staining of samples is not required, means that samples can be easily recovered after analysis and subsequently used for other tests or measurements. Recording a spectrum in ATR mode is usually rapid (second – minutes) and there is a quick sample turnaround between data acquisition. However, a consequence of the design and physics of ATR is that only the first few microns into the surface of samples can be detected. This can be an advantage when studying samples with multiple layers if only the top layer is of interest; for example tissue samples where spectra of only the epithelial layer is desired. However, problems can arise with liquid samples, such as blood and urine, because very low concentration components may not be detected. Low concentration can be overcome by drying samples directly on to the prism to increase their concentrations and, therefore, their signal intensities. Selective partitioning of samples can also occur due to the hydrophobic nature of the ATR prism, leading to over- or underestimation of sample components that either concentrate at the prism surface or are excluded from the active area.

An additional factor to consider with ATR experiments is that the pathlength, and therefore absorbance, is frequency dependent because the depth of penetration of the evanescent wave is dependent on the frequency of the incident light. It is also dependent on refractive index, which can introduce error when comparing samples with different refractive indices.

Before the advent of ATR devices, IR experiments were generally run in transmission mode. In general for these experiments, pathlength must be very short to enable enough light to get to the detector so samples must be thinly sectioned and/or sandwiched between two IR-transmitting windows. Therefore, sample preparation for transmission experiments is inherently more labour intensive than for ATR experiments and it is not always possible to recover samples for other measurements. In addition, the sample chamber must be purged or evacuated in order to remove water vapour from the optical path and this increases the turnaround time of each sample.

In transmission mode the signal size is not dependent on the frequency of the

incident light or the refractive index of the sample, so it can be used to study and compare samples with varying refractive indices, but variations in sample thickness (hence pathlength) do affect signal size and should be taken into account. Unlike in ATR experiments, the entire thickness of the sample is probed allowing components throughout the sample to be investigated, not just those at the very surface. Tissue sections studied in transmission mode are typically 5-10 μm thickness to optimise signal size whilst allowing enough light to get to the detector. Longer pathlengths are possible for less absorbing materials. The transmission mode has a major advantage when applied to vapour because the pathlength can be increased to tens of centimetres or even metres with the use of enclosed chambers and mirrors in the beam path, hence greatly increasing detection sensitivity.

1.1.4 Line shapes in infrared spectra

When a molecule absorbs mid-IR energy it makes a transition from its ground state to an excited vibrational state. In the vapour state, such transitions occur at well-defined frequencies that produce very narrow bandwidth bands in IR spectra. However, interaction with other molecules in the environment influences the precise frequency of vibration.

Initially, after excitation, groups of molecules vibrate coherently. The lineshape of the absorbance band depends on the rate of loss of coherence (coherence lifetime, τ_c) and on the rate of relaxation to the ground state (amplitude correlation time, τ_a), which are in turn influenced by inter- and intramolecular interactions. When τ_a is greater than τ_c a Lorentzian profile is observed (Figure 1.4 A); this is the typical line shape for gases because the rapidly moving particles rotate and collide frequently resulting in loss of coherence. In solids, Gaussian profiles are observed because the molecules are not rotating and colliding (Figure 1.4 B). Liquids exist between these two states leading to IR profiles that combine Lorentzian and Gaussian components [7, 8]. Fitting a curve to an IR spectrum allows individual components that contribute to the overall profile to be extracted. This can reduce problems associated with overlapping bands but relies on some knowledge of the expected composition of an absorbance band.

1.1.5 FTIR microspectroscopy

FTIR-microspectroscopy combines chemical imaging with high spatial resolution and can either be mapping or imaging; mapping experiments take individual spectra

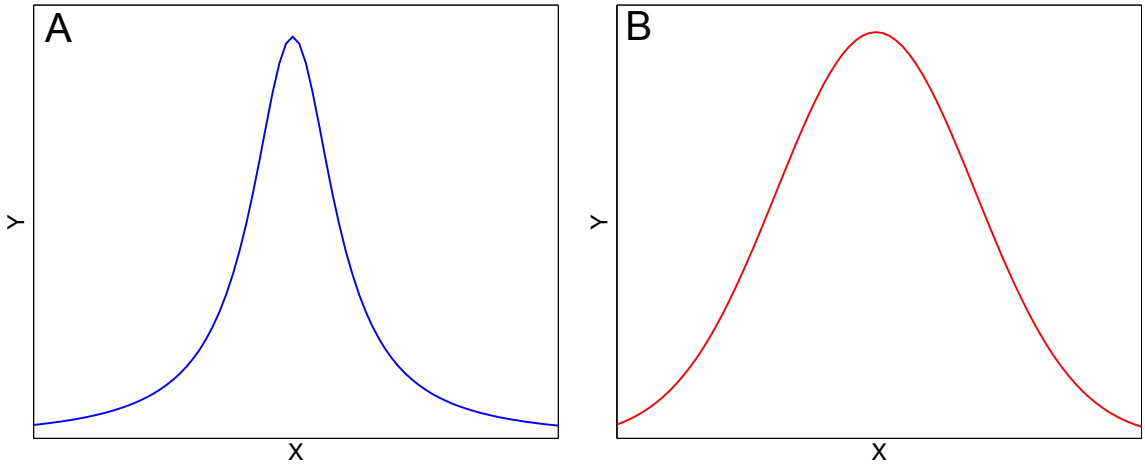


***Figure 1.4: Lorentzian and Gaussian line shapes**. (A) Lorentzian profile typical of gases. (B) Gaussian profile typical of solids.*

sequentially across a defined region of the sample, whilst in imaging experiments a focal plane array (FPA) detector is used to simultaneously record multiple spectra across the imaging area. High spatial resolution of 5 – 10 μm can be achieved, allowing imaging of individual cells or even organelles [5].

FTIR microspectroscopy can be performed in ATR or transmission mode and a synchrotron IR source can be used to significantly enhance the signal-to-noise ratio, although spatial resolution of 3-4 μm has been achieved with a globar source [9]. The ability to generate high quality data using a small and inexpensive globar source is essential if FTIR-microspectroscopy is ever going to be applied to a clinical setting. False colour maps can be generated from FTIR-imaging data that represent the distribution of a band of interest across the sample. Individual spectra from points of interest can also be extracted for analysis.

FTIR microspectroscopy has been used in many areas of research, the earliest applications being the analysis of polymer materials [10], for example, in identifying layers in multilayer films used in packaging [11], and for following degradation of polyolefin films [12]. Other applications have included structure and composition analysis for conservation studies of articles of historical interest, such as paintings [13]; determination of cell wall composition and architecture of wheat endosperm and how it is related to grain hardness [14]; and monitoring of biofilms for bacterial activity and responses to environmental changes [15].

FTIR-microspectroscopy is increasingly being applied to biomedical science. For example, FTIR microspectroscopy has been used to follow changes in protein secondary structure as a means of studying protein aggregation associated with diseases

such as Parkinson's, Huntington's and Alzheimer's [16]. Kazarian and co-workers used the technology to study polymer/drug formulations to establish mechanisms of drug release for optimising drug delivery methods [17]. Perhaps the fastest growing area of FTIR microspectroscopy research is in cancer diagnostics. Work includes grading of tumour samples (molecular histopathology) [18, 19]; distinguishing cancer types [19, 20]; metabolic and biochemical profiling of tumours [21, 22]; identifying cell types [23]; assessing response to chemotherapy [24]; detection of stem cells in the epidermis [25] and stem cell-like clusters in esophageal cancer cell lines [26]; and studying cell division and motility in cancer cells [27].

1.2 Raman spectroscopy

An alternative and complementary form of vibrational spectroscopy to IR is Raman. Raman spectroscopy is based on scattering, rather than absorption, of photons as is the case for IR spectroscopy. Raman scattering is inelastic scattering. Incident light interacts with molecules causing transitions to higher virtual energy levels. In non-resonant Raman, the frequency of the excitation light is such that a molecule is excited to an unstable virtual energy level and quickly re-emits the photon. Resonant Raman is a special case in which the excitation frequency is much closer to that of a molecular electronic transition. Relaxation occurs by emission of light of the same frequency as the incident light (elastic scattering), or of higher or lower frequency (inelastic scattering). If the light emitted is of a lower frequency the molecule returns to a higher energy level than the initial state, this is Stokes scattering. Anti-Stokes scattering occurs when the emitted light is higher frequency than the incident light because the molecule returns to a lower energy level than the initial state. Most molecular vibrations exist in the ground state under ambient conditions so anti-Stokes scattering is weaker. The Stokes Raman spectrum is the most commonly studied.

Although the basis of Raman spectroscopy has similarities to IR spectroscopy, the information produced is complementary due to different selection rules. This often results in normal vibrations that are strong in Raman being weak in IR, and vice versa [2].

Incident light in Raman spectroscopy is typically laser light in the visible, UV or near-IR region. The use of near-IR wavelengths reduces the inherent fluorescence that is usually present that can mask the Raman signals of interest [28].

A major advantage of Raman in biology is that water is a poor scatterer and does not have a strong signal in Raman spectra, in contrast to its strong bands in IR

spectra. However, inelastic light scattering in non-resonant Raman is a rare event so Raman signals are very weak in comparison to absorbance signals [29]. Resonant Raman generally produces stronger signals than non-resonant Raman, but the interference from fluorescence becomes more of a problem.

Raman spectroscopy has been used for numerous biomedical applications including; discriminating between tumour types [30–33]; characterising molecular composition of lung tissue for identification of lung malformations [34]; discriminating between normal and malignant ovarian tissue [35]; diagnosing breast cancer [36] and lung cancer [37]; identifying neoplasia in Barrett's oesophagus [38]; and cancer grading and molecular histopathology [39, 40].

1.3 Data analysis methods

1.3.1 Feature extraction

There are a number of approaches for analysing FTIR data and the method used largely depends on the nature of the data and what information is required. The simplest methods involve specific peak analysis, such as correlating peak positions with reference spectra, comparing peak intensities, integrating peak areas and measuring peak shifts. These processes can be applied to absolute or second derivative spectra.

Peak areas can be used to quantitate specific components and can be determined through integration of the recorded spectrum or by curve fitting and integration of the simulated curve; the latter method improves the specificity of the analysis. Single peaks may be analysed but correlating multiple peaks or points on the spectrum to a compound of interest is more reliable because many compounds have overlapping bands. More complex multivariate analysis methods for mixed samples are also available [41, 42]. Linear and non-linear multivariate regression can take into account the whole spectral range or a selected subsection to generate a linear predictive model from calibration spectra that can be used to quantitate unknown sample data. However, these methods require that the number of observations (i.e. sample number) is greater than the number of variables (i.e. absorbance intensity at each wavenumber in the case of IR spectra) and, since this is often not the case in FTIR spectroscopy, a further processing step to computationally reduce the dimensionality of the data can be applied. Principal component analysis (PCA) and partial least squares (PLS) regression are two commonly used methods. In PCA, a new set of linearly uncorrelated variables, called principal components (PCs), are generated from the original dataset by projecting the data to a new

axis. The variables in the original dataset are recombined so the new PCs are orthogonal to one another (this reduces redundancy in the data and hence reduces dimensionality) and capture the maximum variation in the data. PC1 describes the maximum variance in one axis direction in the data, PC2 describes the maximum variance in an orthogonal direction and so on. The first few PCs typically describe 90% or more of the variation in the data and can be used in subsequent analyses. Inclusion of PCs beyond the first few in these analyses often does not improve performance, but if they are to be used the PC loadings plots, which describe which features in the data contribute to each PC, should first be carefully examined to ensure they do not simply represent differences in noise. PLS works on a similar principle as PCA except that the new components, called latent variables (LVs), are formed so that they both maximise the variance in x (e.g. the absorbance intensity at each wavenumber) whilst maximising the covariance with y (e.g. the disease class or concentration). The LVs generated by PLS contain more information than the PCs generated by PCA about the relationship between the variables (y) and observations (x). PCA or PLS can be combined with multivariate regression to build calibration models for IR spectra [43–45].

1.3.2 Unsupervised and supervised classification methods

Alternative analytical techniques can be used to classify spectra based on spectral pattern recognition. These multivariate analysis methods can be unsupervised, in that they do not use any prior knowledge about the classification of the spectra (as is the case for PCA described in the previous section), or supervised, in that they use knowledge of group membership (PLS regression is a supervised technique) [41, 43].

Clustering methods including hierarchical cluster analysis (HCA) and k-means clustering are common examples of unsupervised classification. Clustering analyses measure the similarities between observations (e.g. spectra) by defining a distance measure (e.g. Euclidean distance or Mahalanobis distance) and calculating pairwise the distance between observations. Spectra can then be sorted into a defined number of clusters according to their 'closeness' to other spectra. This allows underlying structure in a dataset to be identified and is an unbiased means of determining whether two spectra from the same group are more closely related to each other than to the spectra of other group members [41–43].

Supervised classification involves generating a predictive model using a known training dataset which can then be tested on an unknown dataset. Discriminant analysis (DA), linear DA (LDA) or PLS-LDA, multinomial logistic regression (MLR), PLS lo-

gistic regression, support vector machine (SVM), artificial neural networks (ANN) and Bayesian inference are all commonly used supervised classification methods. In PLS logistic regression the LVs generated by PLS regression are used to build a classification model (binary or multinomial) using logistic regression that predicts the probability of a case being positive (i.e. the probability of a sample belonging to a given class) based on the values of the independent variables (LVs in this case) [43].

The performance of classification models is often evaluated by calculating the sensitivity and specificity for predicting membership to each class, and the overall accuracy. Sensitivity describes the models ability to correctly identify positives in a given group whilst specificity describes the models ability to exclude all other samples from that group [46]. Sensitivity, specificity and accuracy are defined in Methods.

1.4 FTIR spectroscopy applied to medical diagnostics

FTIR spectroscopy and imaging have been applied in medical diagnostic research for many years and the literature is expanding rapidly [47]. The technology has been applied to disease biomarker detection in body fluids such as urine and blood. For example, glucose and cholesterol have been measured in whole blood and serum using IR spectroscopy, offering simple alternatives for monitoring patients with diabetes [48,49]; similarly, glucose and lactic acid, as potential cancer biomarkers, have been measured in plasma samples [50]; protein and urea have also been measured in blood, and uric acid, phosphate and creatinine have been measured in urine [49]. FTIR spectroscopy has been applied in reproductive biology to assess oocyte quality and in biomarker detection in synovial fluid to diagnose arthritis [51]. Renal stone composition has also been determined using FTIR spectroscopy [51–53]. Cancer diagnostics is a particularly active area with publications including cervical [18, 54, 55], lung [19, 56], prostate [57, 58], colon [59–62], oesophageal [63–66], gastric [21], brain [31, 67] and skin [55] cancers. FTIR microspectroscopy has also been applied to the detection of stem cells in cancer research [25, 26, 68] and to the characterisation of cell-cycle variations [69].

1.4.1 Biofluids

Biofluids, including urine, saliva, whole blood, plasma and serum, and cerebrospinal fluid (CSF), are an attractive biological sample type for clinical diagnostics because of the relative ease with which they can be obtained. Infrared spectroscopy offers a reagent-free means of quantifying known components in biofluids and for iden-

tification global compositional changes associated with disease (e.g. cancer). In contrast to many traditional biochemical tests, multiple components can potentially be simultaneously detected and quantified from a single spectrum. However, when developing new clinical analytical methods, the new method must be validated by comparison to the current 'gold-standard' and must be tested against a set of samples that spans the physiological range taking into account any possible interferences. Both simple univariate analysis protocols, such as peak integration, as well as more complex multivariate methods, such as PLS regression and deconvolution by peak fitting, have been applied to analyses of biofluid spectra.

1.4.1.1 IR analyses of blood and blood products

Whole blood, serum and plasma have been analysed by a number of groups using IR spectroscopy for detection and quantitation of proteins, sugars and carbohydrates, lipids, creatinine and metabolites as well as disease-specific biomarkers. Pattern recognition techniques have also been applied in an attempt to discriminate between diseased and healthy samples.

Several groups have combined transmission IR spectroscopy with multivariate calibration methods to detect and quantitate glucose, triglycerides, urea, cholesterol, albumin, lactic acid and total protein in whole blood, serum and plasma samples, and achieved high correlation with reference methods [48, 50, 70, 71]. Quantitation of glucose has frequently been achieved using the ~1200-900 cm^{-1} spectral region, with specific peaks identified at 1035/1033, 1078, 1104 and 1146 cm^{-1} [48, 71]. Budínová *et al.* (1997) [48] identified spectral peaks at 955, 1057 and 1371 cm^{-1} as being associated with cholesterol, whilst lactic acid is reported to contribute to absorbance at 1127 cm^{-1} according to Petibois *et al.* (2006) [67].

ATR-FTIR spectroscopy has also been applied to the analysis of blood products. For example, Heise *et al.* (1995) [72] quantified glucose from ATR-FTIR spectra of spiked plasma samples. Hoşafçi *et al.* (2007) [49] analysed ATR-FTIR spectra of whole blood and plasma samples for albumin, cholesterol, glucose, total protein, urea and triglycerides. PLS regression was used to build calibration models for each analyte in the three sample types. Good correlation between ATR-FTIR and clinical methods was obtained for all analytes, although the precise regions of the spectra used for calibration in each case were not defined. The method had good sensitivity for quantitation of some components, such as urea, protein and triglycerides, but for others, thresholds were defined, below which quantitation of the given component was inaccurate but detection was still

possible.

Detection of specific compounds in blood samples can be used diagnostically for known diseases. For example, Petrich *et al.* (2000) [73] analysed transmission spectra of dried serum samples from healthy and diabetic patients. PLS analysis was used to quantitate glucose, triglycerides and cholesterol from the IR spectra and good correlation with the gold-standard analytical methods (correlation coefficients r^2 = > 0.9) was achieved. A supervised discriminant analysis with leave-one-out cross validation or bootstrapping resampling (where data are randomly assigned to training and test sets) methods gave good sensitivity, ranging from 71 to 92 %, and specificity, ranging from 73 to 100%.

Another study (Lacombe *et al.* (2015) [74]) used the 1200-900 cm^{-1} region of transmission IR spectra of dried blood plasma samples to diagnose galactosemia, a metabolic disorder in which galactose accumulates in the blood and urine. The galactosemia patients spectra had lower overall absorbance in this region compared to the healthy group and to a diabetic patient group. The diabetic patient spectra had the highest overall absorbance in the same region. Absorbances in this region were attributed to carbohydrate but no explanation was given to explain why the galactosemic group showed the lowest overall absorbance despite the expectation that galactosemic blood would contain higher concentrations of galactose than healthy blood. It may be that the overall carbohydrate profile in galactosemic patients is different to healthy patients, or that galactose does not contribute as strongly in this region. Comparison to model spectra of galactose would have been useful to help determine the cause of this discrepancy. SVM classification with leave-one-out cross validation was used to classify the spectra. The model performed best for separating diabetic and galactosemic patients spectra where sensitivity was 93-95% with 97-100% specificity, but the accuracy for discriminating between healthy and galactosemic patients spectra was lower. However, no loadings plots were presented to show which specific spectral features were contributing to the separation of the groups.

Other diseases studied using transmission IR spectroscopy of serum samples include liver fibrosis [75] and bovine spongiform encephalopathy (BSE) in cows [76]. The cow serum samples were measured both wet and dried, with dried samples affording the best sensitivity and specificity. This study demonstrated that there is a compromise to be made between the ease of sample preparation and reproducibility afforded by liquid sample analyses, and the lower sample volume requirements and improved accuracy of dried sample analyses [76]. Since the limitation in measuring the liquid samples, in this

case, was due to the material used in the optical windows (CaF_2 rather than the ZnSe used for dried samples) there is potential for further optimisation, using transmission or even ATR sampling methods, to improve the model performance when using liquid samples.

Methods of detecting several cancer types in ATR-FTIR spectra of blood and plasma samples have been developed, including bladder cancer [77], glioma [78], ovarian and endometrial cancers [79]. Ollesch *et al.* (2013) [77] identified problems with reproducibility when drying plasma and serum samples, including the need for very precise volume measurements and samples drying in uneven layers, that resulted in spectral artifacts that would make clinical translation difficult. They developed a robotic dispensing system to overcome these problems and were able to obtain 93 % sensitivity for detecting bladder cancer, but at only 46% specificity.

ATR-FTIR spectra of serum samples has also been applied to the diagnosis of renal failure by Khanmohammadi *et al.* (2013) [80]. Samples were recorded unprocessed and wet and analysed using a supervised pattern recognition method using soft independent modeling of class analogy (SIMCA). Spectral differences at 1470, 1370 and 1320 cm^{-1} were identified between normal and renal failure patients. A spectral window encompassing these wavenumbers was used to build the classification model which performed with 100% sensitivity and 91.3% specificity.

1.4.1.2 IR analyses of urine

Urine can be collected non-invasively and without the need for a trained professional to be present. It can also be collected frequently and stored for several days. It is therefore ideal as a diagnostic medium if it can provide clinically useful information.

Like blood, urine has been studied with IR spectroscopy and methods have been developed to detect and quantitate specific urinary components from IR spectra. Heise *et al.* (2001) [81] used ATR-FTIR spectroscopy to analyse urine samples for urea, creatinine, uric acid, glucose, total protein, phosphate and sulphate. Spectroscopic determination of pH was also achieved by monitoring the equilibrium of mono- and di-basic phosphate, which differ in their IR spectra. PLS analysis of the following spectral regions was used for calibration: 1600-1000 cm^{-1} for urea, creatinine and uric acid; 1200-950 cm^{-1} for glucose; and 1600 - 1250 cm^{-1} for protein. Correlation coefficients for the ATR-FTIR method versus the reference methods were generally high, ranging from r^2= 0.934 for pH, to r^2= 0.996 for sulphate, although calibration was not possible for glucose and protein.

Similarly, in the study by Hoşafçi *et al.* (2007) [49], mentioned in the previous

section, 100 urine samples were analysed by ATR-FTIR spectroscopy for urea, uric acid, phosphate and creatinine concentrations. Calibration models for additional urine components were built using spectra of artificial urine samples containing urea, phosphate, creatinine, ammonia, sulphate, hydroxybutyric acid, acetoacetate, acetone, glucose and uric acid. Again, good correlation with the reference methods was achieved.

Disease biomarkers have also been detected in urine using ATR-FTIR spectroscopy. Markus *et al.* (2001) [82] combined the technology with PLS regression to analyse urine samples for diagnosis of alkaptonuria, a metabolic disorder in which high levels of homogenistic acid are excreted into the urine. Urine samples spiked with homogenistic acid were used to build and test the model before it was tested against urine from 4 alcaptonuria sufferers. The 1530-1150 cm^{-1} spectral region was found to be optimum for building the model. High correlation of the ATR-FTIR calibration model with clinical results was obtained for the validation samples (r^2= 0.998). Homogenistic acid was also successfully detected and quantified in the alcaptonuria patient urine, although no comparison was made to the reference method. A point to consider is that the model was built using 7 PLS factors but 98% of the variability in the calibration samples was reportedly described by only the first three. Noise may contribute to the remaining 4 factors and the model discrimination may be partly based on noise rather than clinically relevant differences in the spectra.

Many of the studies described here have used leave-one-out cross validation when assessing the performance of the classification models. In most cases, replicas of the test samples were included in the training set, which may lead to exaggerated model performance.

1.4.1.3 Enhancing sensitivity of IR measurements of biofluids

Laminar fluid diffusion interface (LFDI) microfluidic devices have been proposed as a preprocessing technique to enhance the sensitivity of FTIR measurements of biofluids. This technique can be used to separate components in samples based on their molecular weight. It is particularly useful for analysing samples where high concentration components, such as proteins in serum samples, dominate the IR spectrum making it difficult to detect more minor components. In the LFDI device a sample stream and a receiver (e.g. water) stream run in parallel in laminar flow so that there is diffusion between the streams without mixing. Low molecular weight compounds diffuse more quickly into the receiver stream than larger molecular weight compounds such as proteins, so that it becomes enriched with low molecular weight components. The resultant

receiver stream can then be analysed in much the same way as the original sample in transmission or ATR mode, wet or dried.

Shaw *et al.* (2009) [83] tested the use of an LFDI device to improve the sensitivity of detecting creatinine in serum samples. After centrifugation, dried films of either 2 μL unprocessed plasma, or 40 or 120 μL LFDI receiver streams, were measured in transmission mode. PLS calibration models were built using the second derivative transformations of the IR spectra. The LFDI-processed samples showed a 17-fold increase in the creatinine to albumin ratio compared to the unprocessed plasma samples. Absorbance signals from creatinine could then be enhanced by increasing the thickness of the sample and hence increasing the effective pathlength of the IR light. There was also a more than 4-fold improvement in the standard error of calibration for the 120 μL LFDI-processed sample compared to the 2 μL unprocessed sample. However, the separation and drying procedures followed here are laborious and time-consuming so not suitable for clinical translation, and the use of LFDI technology to detect and quantitate other diagnostic compounds in biofluids has not yet been demonstrated. Nevertheless, with further optimisation and integration with ATR technology the LFDI system could provide an effective method of increasing the sensitivity of IR measurements for point-of-care diagnostics.

1.4.1.4 IR analyses of kidney stones

FTIR spectroscopy has proved a useful method for compositional analysis of kidney stones.

Channa *et al.* (2007) [52] recorded transmission FTIR spectra of 8 commonly occurring stone compounds as references for analyses of 58 renal stones. 83% of stone types were either mixed calcium oxalate and uric acid or pure calcium oxalate. Struvite, pure uric acid, and mixed calcium oxalate and aspartate stones were also observed. Cystine was not detected in any of the stones studied.

In stones from a Polish population studied by transmission FTIR spectroscopy by Paluszkiewicz *et al.* (1988) [53] a high proportion were pure calcium oxalate, or a mixture of calcium oxalate and phosphate. Struvite, calcium phosphate, uric acid and a small proportion of cystine stones were also detected.

Mulready and McGoldrick (2012) [84] compiled a library of ATR-FTIR spectra of kidney stones that combined standard reference spectra from common kidney stone components (including calcium oxalate monohydrate, calcium oxalate dehydrate, calcium hydrogen phosphate, hydroxyapatite, carbapatite, magnesium ammonium phos-

phate and uric acid, as well as some rarer components including cystine and 2,8- dihydroxyadenine) with spectra from 'real' kidney stones from patients in the Republic of Ireland. The use of real stones in the construction of the library allowed minor peak shape and position variations, which are dependent on composition and environment, to be taken into account. It was found that FTIR-based methods had superior sensitivity to wet chemical methods and were able to detect minor or rare stone components more readily. However, for some compounds, such as magnesium ammonium phosphate, additional chemical tests were needed for accurate detection.

1.4.2 Tissue analyses and cancer diagnostics

FTIR spectroscopy can potentially detect changes in the biochemical composition of tissues during the progression from normal healthy tissue to invasive cancer. In combination with multivariate analyses, spectral patterns that distinguish cancerous from healthy tissue can be identified; this has become known as spectral histopathology (SHP). Alternatively, analyses of specific wavenumbers can identify associated biochemical changes indicative of disease, for instance glycogen and DNA are repeatedly reported to be altered in cancerous samples and such changes have been detected using FTIR spectroscopy.

1.4.2.1 Cervical and prostate cancers

Mordechai *et al.* (2004) [55] and Mark *et al.* (2004) [18] reported that spectra of sectioned cervical cancer tissue samples imaged in transmission mode had lower absorbance at wavelengths associated with carbohydrate/glycogen, and higher RNA/DNA peak intensity ratios at 1121/1020 cm^{-1} than spectra of normal cervical tissue. Mark *et al.* (2004) [18] applied a probabilistic neural network (PNN) to their cervical cancer dataset and achieved correct diagnoses for 90.95 % of normal, 84.28 % of mildly dysplastic, 89.46 % of moderately dysplastic and 83.26 % of severely dysplastic cervical tissue.

A diagnostic model using transmission FTIR microspectroscopy and principal component discriminant function analysis was built for prostate tissue and scored 92.3% overall sensitivity and 98.9 % overall specificity for classifying tissues into a three band grading system [57]. The same group later applied their methods to evaluate 40 prostate biopsies from 39 men. Importantly, when building their supervised classification model patients included in the training set were not included in the test set. The resultant classifications based on the same three band grading system gave lower sensitivities of 70-84% and specificities of 72-88% than were previously reported by the group, probably

as a result of the independent training and test sets used [85].

1.4.2.2 Lymph nodes

Bird *et al.* (2009) [86] analysed second derivative spectra extracted from transflection FTIR images of lymph node tissue. They combined HCA with Ward's clustering algorithm of the 'phosphate region' between 1350 and 900 cm^{-1} using 9 clusters to reveal the location of micro-metastases in lymph nodes.

1.4.2.3 Barrett's oesophagus

Several groups have applied FTIR spectroscopy to the detection of Barrett's oesophagus and pre-malignant stages of oesophageal cancer. Wang *et al.* (2007) [66] applied LDA with leave-one-out cross validation to ATR-FTIR tissue spectra to distinguish normal squamous tissue from Barrett's oesophagus and gastric tissues, achieving 100 % sensitivity and 97 % sensitivity. Samples containing Barrett's dysplasia showed elevated DNA absorbance bands at 1026, 1081 and 1154 cm^{-1} compared to non-dysplastic Barrett's. A unique band at 1117 cm^{-1} in gastric tissue was attributed to increased glycoprotein. Likewise, Quaroni and colleagues (2009) [65] combined HCA and Ward's clustering algorithm of the 1300-930 cm^{-1} spectral region using 5 clusters to discriminate between normal squamous tissue, Barrett's oesophagus and oesophageal adenocarcinoma from transflection FTIR images. Bands at 1029, 1079 and 1150 cm^{-1} in second derivative spectra were attributed to accumulation of glycogen in Barrett's tissue compared to normal and adenocarcinoma tissue. Barrett's tissue was also associated with absorbance bands at 1080, 1124, 1171 and 970 cm^{-1} which were attributed to mucin.

1.4.2.4 Lung cancer

Lung cancer has also been studied by transmission FTIR spectroscopy, ATR-FTIR spectroscopy and FTIR microspectroscopy. Yano *et al.* (1996) [87] attributed elevated absorbance of the 1045 cm^{-1} peak in transmission spectra of lung carcinoma samples to increased glycogen. They later used FTIR imaging to study lung tissue cryosections and proposed a decrease in the 1467 cm^{-1} band of cancerous samples compared to normal could be due to reduced cholesterol levels [20]. A study of FTIR images of lung sputum recorded in reflectance mode also suggested elevated glycogen in cancer samples based on increased absorbances at 1024 cm^{-1} and 1049/1051 cm^{-1} (normal/cancer). The study also found amide I and II peaks were increased in cancer samples versus normal [56]. Similarly, increases in intensity of the 970 cm^{-1} and 1085 cm^{-1} bands in the

reflectance FTIR spectra of lung cancer cell lines were attributed to changes in DNA structure or concentration [88]. Bird and colleagues (2012) [19] demonstrated the use of spectral histopathology (SHP) as a tool for diagnosing lung cancer using a tissue microarray of 10 normal tissue spots and 70 lung disease tissue spots of different types. FTIR images of the tissue spots were recorded in reflection mode. An artificial neural network was used to develop a binary classification for diagnosing the different tissue types. Normal tissue spectra could be distinguished from all other samples with high sensitivity (99.3 %) and specificity (94.4 %). Sensitivities and specificities of over 90 % were also reported for discriminating between most of the lung cancer tissue types studied. ATR-FTIR spectra were analysed by Sun et al. (2013) [89] who reported significant differences between healthy and lung cancer samples for band widths at ~1303 and ~1240 cm^{-1}, and for peak intensities at 1120 and 1546 cm^{-1}. These features achieved 96.7 % sensitivity.

1.4.2.5 Breast and ovarian cancers

Walsh *et al.* (2012) [90] studied breast tissue microarrays using ATR imaging and Bayesian classification for structural characterisation as an alternative to traditional H&E histology. They were able to identify newly formed blood vessels in the tumour tissue and distinguish these from surrounding stromal cells. They also reported distinctions between myoepithelial cells, which acts as markers for malignancy, and normal epithelial cells, with spectral differences occurring at 3800-2700 cm^{-1}. However, since this region can be influenced by small changes in hydration and because these spectra were normalised to the amide I band, where water also absorbs, changes may also be influenced by the extent of hydration of the dried tissue sections. Finally, terminal ductal lobular units, where many breast cancers originate, were identified in ATR-images by increased absorbance at 1300-1200 cm^{-1} compared to epithelial cells. This was attributed to increased levels of collagen in intralobular stroma, with only low levels of collagen in epithelial cells.

Normal, benign and malignant ovarian tissue transmission FTIR images were studied by Krishna *et al.* (2007) [35]. Malignant tissue spectra had additional absorbance bands at 828 an 857 cm^{-1}, an increase in absorbance between 1200 and 1000 cm^{-1} and an increased absorbance at 1734 cm^{-1}, attributed to lipid. Normal tissue was found to contain higher amounts of protein as judged by positive amide I and II peaks at 1652 cm^{-1} and 1553 cm^{-1} in difference spectra. Benign tissue was found to have increased DNA and lipid content, indicated by positive peaks at 1035, 1086 and 1131 cm^{-1}, whilst

positive peaks in malignant tissue spectra suggested increased lipid (1738 cm^{-1}), DNA (1092 and 1248 cm^{-1}) and glycogen (1045 and 1074 cm^{-1}).

1.4.2.6 Brain cancers

Gajjar *et al.* (2013) [31] applied ATR-FTIR spectroscopy to the discrimination of different brain tumour types (meningioma, glioma including 3 subtypes, and brain metastasis) using linear discrimination analysis (LDA) or PCA-LDA. Lipid (1740 cm^{-1}) to protein (1400 cm^{-1}) ratios in were reportedly higher in normal tissue than in tumour tissue. Phosphate (1045 cm^{-1}) to carbohydrate (1545 cm^{-1}) was reduced in high-grade gliomas, whereas RNA (1121 cm^{-1}) to DNA (1020 cm^{-1}) ratios were reduced in meningioma and slightly reduced in metastatic tumours. Absorbance bands contributing to the discrimination between normal and different brain tumour type spectra were tentatively assigned to glycogen (1018 cm^{-1}), carbohydrate (1173 cm^{-1}), amide I (1543, 1620, 1628, 1686 cm^{-1}), amide II (1582, 1504, 1551 cm^{-1}), lipids (1740, 1470, 1701 cm^{-1}) and phosphate backbone (1103, 1234 cm^{-1}). However, these are speculative assignments as without multiple peak correlations to model spectra it is impossible to assign these changes to specific biochemical constituents. Many of these assignments overlap and are even in contradiction to earlier assignments made when looking at absorbance ratios. Furthermore, the results were not quantified in terms of sensitivity and specificity so it is difficult to evaluate the efficacy of their method.

1.5 Aims

Although the research described in the preceding sections demonstrates the potential of IR spectroscopy in diagnosing a variety of diseases, there has not always been an adequate focus on developing methods that could easily be employed in the clinic, and/or that provide significant advantages over current diagnostic methods. For example, simple diagnostic tests already exist for diabetes with simple glucose assays, and galactosemia can be effectively screened for in newborns by a range of methods, including colorimetric and chromatographic assays of galactose concentration in blood and urine, enzyme activity assays, and genetic testing of blood samples [73, 74]. So, there may not be a compelling need to develop FTIR-based diagnostic methods for such diseases. Conversely, cancer diagnosis remains expensive and time-consuming and so there is a need for improved methods of detection and staging. FTIR methods have been applied to cancer diagnostics, though these methods also are generally complex and

time-consuming, and so are difficult to translate to a clinical setting unless they can be simplified.

FTIR-methods that are able to detect several compounds simultaneously in a given sample may offer advantages even if simple clinical tests for those individual compounds are available. For example, Hosafci (2007) [49], Heise (2001) [81] and Jessen (2014) [70], described the quantitative analysis of multiple components in urine, whole blood and blood plasma. However, these methods have not yet been translated into clinical uses for the diagnosis of specific medical conditions.

The aims of the research presented in this thesis were to develop reproducible IR spectroscopy methods that might be developed into clinical tools for diagnosing diseases for which the current diagnostic methods are lacking in some way. Particular focus was on developing statistically robust methods that are sufficiently simple and inexpensive so that translation into clinical uses could become feasible.

A model compound spectral library is presented at the outset, which was used to help assign features in biological sample spectra presented in subsequent chapters, and can be used to to assess assignments made by others in published work.

Much of the clinical literature describes the analysis of dried samples (e.g. [73], [74], [83], [90], [91]) but the potential differences between wet and dried sample spectra are only acknowledged by a handful of groups (e.g. [77], [76]). Because of this, the influence of hydration state on the infrared spectra of biological compounds, specifically urinary components, was investigated and is described in Chapter 4. These analyses highlighted the dramatic shifts that can occur between IR spectra of wet, dry and partially-hydrated states. They showed the importance of recognition of such hydration effects when interpreting spectra of complex biological samples so that shifts due to difference in hydration are not erroneously interpreted as clinically-relevant signal differences.

Compositional analysis of kidney stones using FTIR spectroscopy has already been incorporated into clinical use (see section 1.4.1.4), but, whilst this is useful in diagnosis, a method that can detect and identify multiple stone constituents in urine prior to stone formation is desirable for better disease management. Kidney stone diseases, especially rare types such as cystinuria, are challenging to diagnose accurately by conventional methods and therefore represent good targets for development of FTIR methods. Chapter 5 describes development of a method for detecting and quantifying cystine, the causative agent of cystinuria, from ATR-FTIR spectra of cystinuric patient urine samples.

The possibility of translation of FTIR spectroscopy in to the clinic is often al-

luded to in the literature but in order to make this a reality the FTIR spectrometer must be simple and stable enough for hospital installation. In addition, the operation protocol must be fast and simple enough to be performed by a non-specialist. Several spectrometer manufacturers now make small, benchtop ATR-FTIR spectrometers with room temperature detectors that are stable to vibrations and easy to use. Chapter 6 describes how one such machine was installed in a clinic at the Royal Free Hospital to analyse urine samples in order to assess whether the data generated were of sufficient quality for diagnostic use. This study also gives an insight into the practicality of routinely operating such a machine in a clinical setting by medical staff with limited FTIR training.

Finally, the single-element detector ATR-FTIR methods used for urine analyses were applied to pancreatic tissue analyses for detection of spectral changes associated with cancer. Much of the FTIR cancer diagnostics literature focuses on using imaging techniques, which, although promising in terms of their ability to detect early biochemical changes associated with disease, are expensive, time-consuming and complex. Since tissue samples are inherently more complex than urine samples they are more challenging to study with simple ATR-FTIR methods, but such methods have a better chance of clinical adoption as, if effective, they offer considerable savings in time, money and resources.

To summarise, the specific aims of this thesis were to:

- Generate a library of absorbance and second derivative ATR-FTIR spectra of wet and dried compounds/materials, in particular focussing on human urine of healthy and diseased patients, and healthy and tumorous human tissue.

- Investigate and explain the effects hydration states have on the infrared spectrum of urea and other urinary compounds, including creatinine, ammonium ions, sulphate, phosphate and cystine.

- Develop a method to detect and quantitate urea and creatinine in urine samples using ATR-FTIR spectroscopy.

- Refine protocols for measuring insoluble material in aqueous samples using ATR-FTIR spectroscopy, particularly cystine, with the aim of developing a new diagnostic method for cystinuria. Extend these urine analyses to other disease biomarkers, including dihydroxyadenine (DHA) for APRT-deficiency, and trimethylamine (TMA) as a marker for trimethylaminuria.

- Establish a large database of patient urine ATR-FTIR spectra and medical case

histories using a simple benchtop ATR-FTIR spectrometer installed in a hospital kidney clinic and correlate with clinical data.

- Analyse infrared signatures of fresh pancreatic tissue resections from normal and pancreatic cancer patients.

Chapter 2

Materials and methods

2.1 Materials

2.1.1 Chemicals

Ammonium chloride, cystine, glucose, hydrochloric acid, potassium hydroxide, potassium phosphate, sodium chloride, sodium dodecyl sulphate and sodium hydroxide were purchased from VWR International BDH. Creatinine, picric acid, potassium sulphate and urea were purchased from Sigma-Aldrich. Sodium phosphate was purchased from Fisons Chemicals.

2.1.2 Clinical samples

Five urine samples were collected from volunteer healthy donors at University College London and Imperial College London, UK, for hydration effect analyses. These were donated anonymously and so age/sex information is not available.

For developing urinary cystine analyses urine samples from 22 cystinuric patients, 5 healthy donors (controls) and 11 patients with other renal conditions (for creatinine analysis only), were collected by Professor Robert Unwin and nurses, Faith Matjiu and Cameron Davey, at the Kidney and Urology Centre, Royal Free Hospital, London, UK. Urine was collected in 30 mL Sterilin specimen containers and measured fresh where possible, a small number of samples were frozen and stored at -20 °C before analysis. All urine samples were collected after obtaining informed consent. The study was approved at Royal Free London NHS Foundation Trust (R&D ref: 7727; REC number: 05/Q0508/6).

For dihydroxyadenine analyses a 2,8-dihydroxyadenine (DHA) stone/crystal sample and additional urine samples were supplied by Kristin Hafsteinsdóttir, Reykjavik, Iceland. Urine samples from two patients, one diagnosed with APRTd and the other with suspected APRTd, were supplied by Professor Robert Unwin, Royal Free Hospital, London. Three additional urine samples from three members of a family of APRTd sufferers were supplied by Emma Miler, Southend University Hospital, UK. Two urine samples from a rat-model of APRTd (adenine-fed) and one urine sample from a control rat were supplied by Richard Fish, Royal Free Hospital, London. All urine samples for DHA analyses were received frozen and stored at -20 °C before analysis.

For analyses of urine samples for kidney-related diseases using a benchtop spectrometer, 295 urine samples from 256 patients were collected by staff nurses, Faith Matjiu and Cameron Davey, at the Kidney and Urology Centre, Royal Free Hospital, London. Forty two additional urine samples from healthy volunteers with no known medical conditions were collected at the Royal Free Hospital, Universtiy College London (UCL) and

Imperial College London. All samples collected at the Royal Free Hospital were measured fresh. Samples collected at UCL and Imperial were stored at 4°C and measured within 24 hours.

Forty-three pairs of matched human pancreas tissue (43 tumour and 43 normal) resections plus 3 unmatched tumour resections and 1 unmatched normal pancreas resection were collected during major pancreatic surgery at Hospital Curry Cabral, Universidade Nova de Lisbon, Portugal, and immediately frozen on dry ice. Each pair of matched tissue samples came from one individual, where the 'normal' pancreas tissue samples were taken from regions of 'healthy' pancreas distant from the site of the tumour. The tissue samples were shipped to UCL, London by Graça Soveral, Lisbon University, via courier on dry ice.

2.2 Spectroscopic methods

2.2.1 ATR-FTIR and transmission FTIR spectroscopy

Spectra were recorded with a Bruker IFS/66S FTIR spectrometer with a globar light source, liquid nitrogen-cooled MCT detector and a KBr beam splitter. The interferometer scanner velocity was set to 40 kHz and the optics were purged with dry air. Experiments in ATR mode were run using a SensIR 3 reflection silicon microprism with ZnSe optics with a 1.5 mm aperture (Figures 2.1 and 1.3). Samples were placed in direct contact with the microprism and the depth of penetration of the evanescent infrared wave was approx. 0.5 - 2 μm depending on the wavelength and the refractive indices of the prism and sample. For transmission mode experiments a 1 mm aperture was used and samples were sandwiched between 2 mm thick CaF_2 windows.

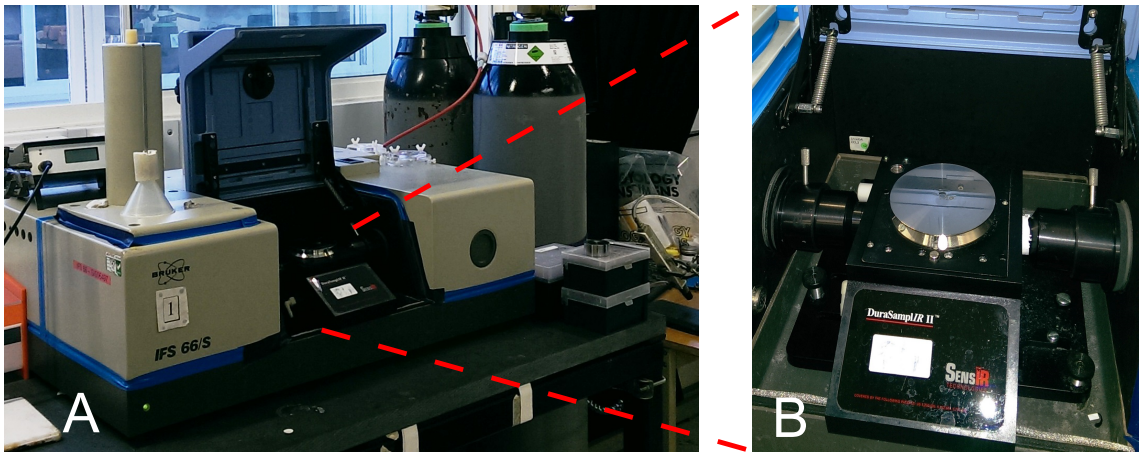


***Figure 2.1: Bruker IFS/66S spectrometer**. (A) The FTIR spectrometer is fitted with (B) a SensIR 3 reflection silicon microprism (also shown as a schematic in Figure 1.3).*

Data were recorded in the mid-infrared range between 4000 – 600 cm^{-1}. In order to reduce noise 500 interferograms at 4 cm^{-1} resolution were coadded before Fourier transformation to generate the power spectrum. The absorbance spectra were computed versus a background of the clean prism surface (1000 interferograms) by application of the Beer-Lambert law. Frequencies given are accurate to approx. 1 cm^{-1}. All samples were at room temperature during analysis.

2.2.2 FTIR spectroscopy using a benchtop FTIR spectrometer in the clinic

Spectra were recorded with a Bruker Alpha FTIR spectrometer (Figure 2.2) with an air-cooled globar source, a room temperature DTGS detector and a multi-layer coating Quarz beam splitter. The Alpha was fitted with a permanently aligned RockSolid interferometer with gold-mirrors. The housing was sealed and water vapour minimised by convective desiccation. Experiments were run in ATR mode using a Platinum ATR single reflection diamond microprism. Data were recorded between 4000 – 600 cm^{-1}. The scanning speed of the Alpha interferometer is inherently slower than that of the IFS/66S. Therefore, 50 interferograms at 4 cm^{-1} resolution were coadded before Fourier transformation to provide a compromise between acquisition time and signal-to-noise ratio of spectra. The absorbance spectra were computed versus a background of the clean prism surface (100 interferograms). The Alpha was programmed to run diagnostic tests to check wavenumber accuracy and water vapour every three months.

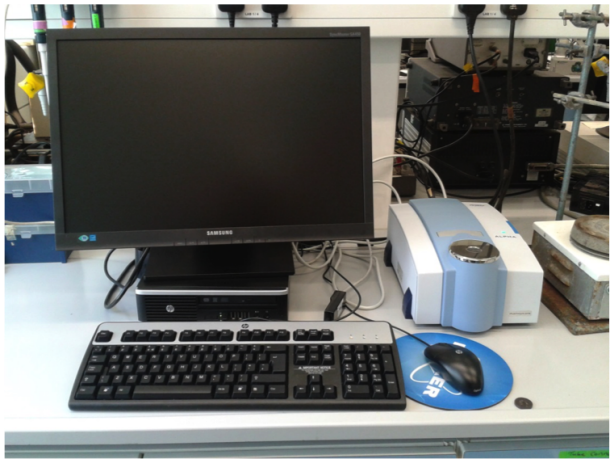


Figure 2.2: Bruker Alpha FTIR spectrometer.

2.2.3 Transmission FTIR spectroscopy of volatiles

Spectra were recorded with a Bruker Vertex 80v FTIR spectrometer with a globar light source, liquid nitrogen-cooled MCT detector and a KBr beamsplitter. The interferometer scanner velocity was set to 40 kHz and the optics were evacuated. Experiments were run in transmission mode with a 3 mm aperture and using a custom-built device designed to contain vapour whilst providing a long (23 cm) optical pathlength through the sample to increase sensitivity. Data were recorded between 4000 - 800 cm^{-1}. 500 interferograms at 1 cm^{-1}resolution were coadded before Fourier transformation and absorbance spectra were computed versus a background of water vapour (500 interferograms).

2.3 Sample handling

2.3.1 ATR-FTIR measurements of liquid samples

All frozen samples were thawed at room temperature before analysis. Samples were mixed by vortexing prior to analysis to ensure homogeneous distribution of solutes throughout the sample. Spectra of solutions and liquids were recorded by pipetting a 20 μL aliquot of a sample on to the ATR microprism and recording FTIR spectra as described. In some cases it was necessary to cover the sample with a small, concave volatiles cover to avoid evaporation of solution constituents during measurement.

2.3.2 Drying samples and dealing with insoluble material for ATR-FTIR spectroscopy

In some cases the signal intensities arising from compounds of interest were much smaller than those of the many other IR-active components. In addition, since only the first few microns of a sample are probed by the evanescent wave of the IR beam in ATR mode, settling of any particulate material onto the prism surface will lead to unstable spectra. To both increase sensitivity and avoid instabilities due to insoluble particle sedimentation, liquid samples were often dried on to the prism surface. After mixing by vortexing, a 1-5 μL aliquot (depending on concentration) of a sample was placed on the prism. Samples were dried by passing a gentle stream of dry nitrogen gas over the sample at a flow rate of 300 mL/min (Figure 2.3). Extent of drying was judged by the absence of water bands in the 3500 – 3100 cm^{-1} and 2100 cm^{-1} spectral regions.

The silicon microprism has hydrophobic properties, often causing more hydrophilic materials to dry unevenly or 'clump' slightly above the surface of the prism. This leads to poor reproducibility and reduced sensitivity. Therefore, once dried, even pressure was applied to the sample using a standard press (from SensIR) to ensure more even and reproducible contact between the samples and the surface of the ATR prism.

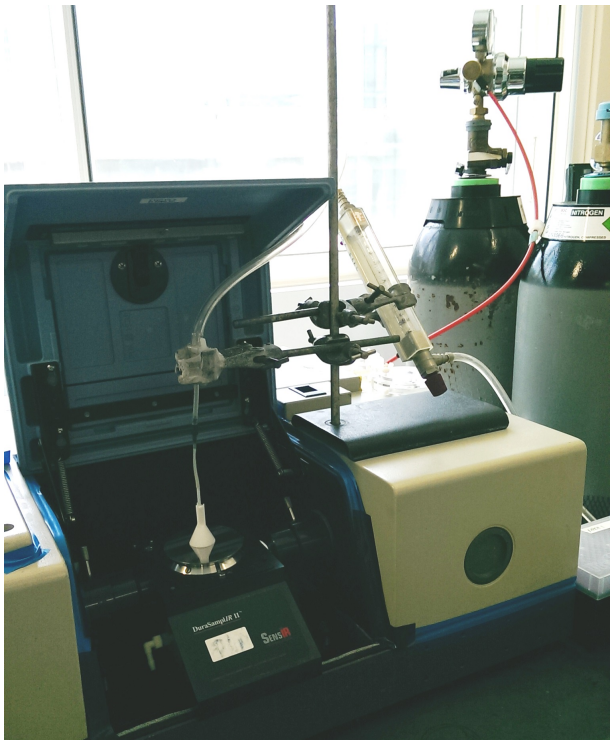


Figure 2.3: Drying a sample on to the ATR prism with a gentle stream of dry nitrogen gas.

2.3.3 Separating insoluble material

The sensitivity of detection of insoluble materials was increased by removing soluble components from the sample before analysis. Typically, 1 mL sample aliquots were centrifuged at 16000 g_{AV} for 5-40 minutes to separate soluble and insoluble fractions. Each pellet was resuspended in 1 mL double distilled water and mixed by vortexing. 5 µL aliquots of these suspensions were dried onto the prism surface and their spectra recorded as described.

2.3.4 ATR-FTIR measurements of solids

Solid materials were ground into a fine powder before analysis. A small amount of the powder was placed directly on to the ATR prism and even pressure was applied using a standard press from SensIR to ensure good contact between the powder and the prism surface.

2.3.5 Monitoring spectral changes associated with hydration states

2.3.5.1 Controlled rehydration of urea

3 µL of a 50 mM urea solution, prepared in double distilled water, was dried onto the crystal with a gentle stream of dry nitrogen gas at a flow rate of 300 mL/min. This led to loss of liquid water bands in the 3500–3100 and 1700-1600 cm^{-1} regions. Drying was continued until the absorbance spectrum was stable in the 1700-1400 cm^{-1} region. The sample was then allowed to equilibrate for several hours in atmospheres of controlled humidities, imposed by placing different saturated salt solutions [92] (Table 2.1) in a well within a chamber that enclosed the sample (Figure 2.4). Spectra were recorded when the 1700-1400 cm^{-1} region had stabilised.

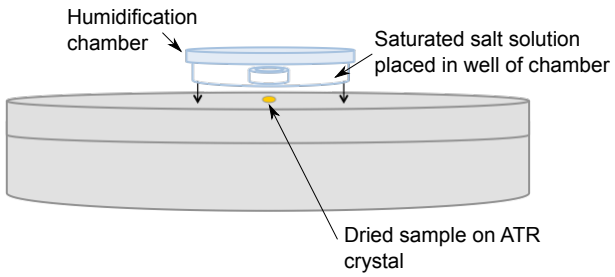


Figure 2.4: Schematic of the humidification chamber device.

Table 2.1: Salt solutions and corresponding humidity at room temperature.

Saturated salt solution	Humidity at room temperature (%)
H_2O (no salt)	(taken as)100
KNO_3	94
KCl	84
NH_4SO_4	81
NH_4Cl	79
NaCl	75
$NaNO_3$	74

2.3.5.2 Serial dehydration of urinary components and urine

Spectral changes associated with varying hydration states were tracked by recording serial spectra during the dehydration process. Three microlitre aliquots of different solutions (50 mM urea, 10 mM creatinine, 29 mM ammonium chloride pH6, 100 mM potassium sulphate pH7, 50 mM potassium phosphate pH7, 25 mM glucose, 12 mM cystine suspension) and urine samples (diluted 1 in 4) were dried onto the prism and rehydrated by placing double distilled water in the humidification chamber described above. Once the spectrum had stabilised, the humidification chamber was removed and a series of spectra (each an average of 10 interferograms; approx. 4 s data acquisition time) were recorded whilst dry nitrogen gas was passed over the sample until the sample had stabilised (typically 50-100 spectra over 3-7 minutes).

2.3.6 ATR-FTIR measurements of tissues

Samples were thawed on ice before their ATR-FTIR spectra were recorded as described. In order to allow for possible heterogeneity of the tissue surfaces, two spectra were recorded per sample and for each spectrum the sample was placed in a different orientation on the prism. The prism was cleaned with double distilled water and ethanol between recording the first and second spectra of each sample. Morphological features such as colour, texture and firmness were noted.

2.3.7 Vapour phase measurements of volatile samples in transmission mode

A custom-built module compatible with the Bruker Vertex 80v spectrometer was built in order to enable measurement of transmission FTIR spectra of volatile samples (Figure 2.5). The device consisted of a stainless steel tube with CaF_2 windows at either end to allow the IR beam to pass through. In the centre of the tube there was a small well on a removal plate in which the sample was placed. This allowed volatiles to evaporate and their vapours to fill the tube through which the IR beam was passed. Once the samples were placed inside the device, it was sealed and left for several minutes at room temperature to allow the vapour and solution to equilibrate before recording the spectra.

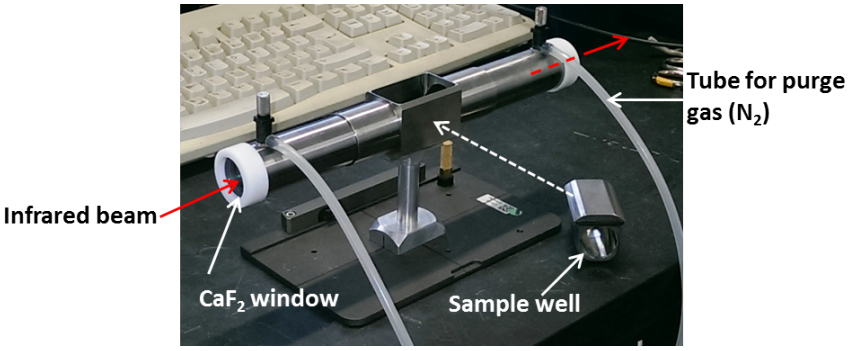


Figure 2.5: Vapour chamber device for transmission measurements of volatiles. *Dry nitrogen gas can be passed through the tube to purge atmosphere before sample insertion.*

2.4 Data processing and analyses

Spectral pre-processing techniques can be used to correct interferences and measurement artefacts, remove noise, simplify spectra for easier deconvolution, and normalise spectra. The following data processing steps were performed in OPUS 7, OriginPro 8.6 or MATLAB R2012b using in-house scripts.

2.4.1 Data preprocessing

A major consideration when recording and analysing IR spectra is the strong absorbance bands of water and water vapour. Since biological materials, and in particular biofluids such as urine, are high in liquid water, it is often necessary to remove the water contribution from IR spectra to make deconvolution of other constituents easier. In cases where samples cannot be dried before measurement, or where drying is undesirable,

pure liquid water spectra can be subtracted from sample spectra (Figure 2.6 B). The water band at approx. 2200 - 1900 cm^{-1} falls in a region of the spectrum that is relatively free from other bands and the loss of this band in the sample spectrum can be used to monitor the extent of subtraction. This can be an effective pre-processing method, although the spectrum of water can be influenced by solutes, particularly strongly hydrogen-bonding compounds, resulting in less accurate subtractions. Where samples have been prepared in other solvents, buffers or media such as paraffin, spectra of these can also be subtracted from sample spectra.

Here, a spectrum of liquid water was subtracted from each wet sample spectrum until the integrated peak area between 2200 - 2000 cm^{-1} was minimised, whilst not distorting the rest of the sample spectrum.

Contributions from water vapour can be partially removed from spectra by fractional subtraction of a pure water vapour spectrum (Figure 2.6 A). However, this can be imprecise as the water vapour spectrum is a distortion of the true spectrum since the very narrow bands cannot be accurately resolved. Therefore it is best to minimise water vapour contributions at the time of measurement. This can be achieved by purging the optics of the spectrometer with dry air/nitrogen gas or desiccants or keeping them under vacuum to remove any internal water vapour. Since environmental water vapour varies and is influenced by weather, humidity and human breath, it is important to record background spectra close in time to sample measurements to minimise the difference in atmospheric water between background and sample scans. In addition, water vapour spectra for subtraction should be recorded on the same instrument and with the same measurement parameters as the sample spectrum with environmental conditions matched as closely as possible [41, 93, 94].

The contributions of water vapour variations in the optical path were removed from all sample spectra by fractional subtraction of a pure water vapour spectrum until further subtraction began to introduce baseline distortions in the 2100 – 1250 cm^{-1} region.

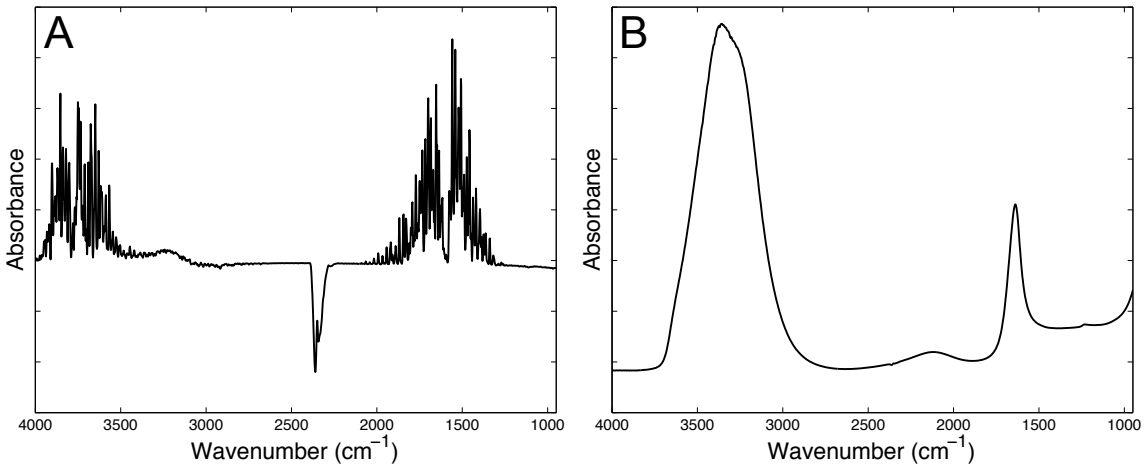


Figure 2.6: Water vapour and liquid water spectra. *(A) Pure water vapour spectrum. A background measurement of the clean prism was recorded. The purge gas to the optics was then turned off and water vapour was allowed to accumulate for 20 minutes before a sample spectrum was recorded of the clean prism. The spectrum is characterised by many sharp absorbance bands between ~2080 and 1290 cm^{-1} and between ~3270 and 4000 cm^{-1}. (B) Liquid water spectrum. Liquid water has a strong broad absorbance band between ~3750 and 3000 cm^{-1} and a prominent sharper band at 1638 cm^{-1}. There is also a broad low intensity absorbance band between 2300 and 1900 cm^{-1}.*

True random noise can be reduced by simply averaging multiple scans of the same sample. Further de-noising of spectra can be achieved through smoothing, commonly using the Savitsky-Golay algorithm, although care must be taken to avoid interpreting distortions as true signal variations and to avoid smoothing out true signals. Other methods of de-noising include wavelet de-noising (WDN), which is thought to be most effective when the sample gives rise to high-frequency absorbance bands. Spectra recorded on the benchtop Alpha spectrometer were smoothed using the Savitsky-Golay algorithm with a 9 point window.

The Savitsky-Golay algorithm is also frequently applied before differentiation of absorbance spectra since differentiation inherently enhances spectral noise. Calculation of the second derivative of the absorbance spectrum narrows peaks allowing easier deconvolution, particularly in spectral regions where multiple absorbance bands overlap, and removes broad baseline drifts. Comparison of derivative spectra with corresponding absorbance spectra can be useful to help identify which bands are noise and which are true signals. Baseline correction techniques, such as piecewise linear and rubber band baseline correction can also be applied to spectra [41, 42]. Second derivatives of the absorbance spectra were computed using the Savitzky-Golay algorithm with a 17 point window (or a 9 point window for pre-smoothed spectra recorded on the Alpha benchtop spectrometer).

Normalisation of spectral data is often required to allow for differences in sample thickness or concentration so that different sample spectra can be directly compared. Spectra can be normalised to the intensity (min-max) or area of a peak expected to be constant between samples (often, amide I or II peaks for tissue samples or a urea peak in urine). In general, it is best to avoid normalisation to a peak where water also absorbs as water content can vary between samples. Vector normalisation is often applied to derivative spectra where specific single peak for normalisation cannot be identified; the data are scaled so that the sum of the square of all intensity values equals one. Standard normal variate (SNV) normalisation is another alternative that can also be used for scatter correction [41, 42, 95, 96].

Here, spectra were typically normalised to the intensity of a dominant band in the spectrum. Tissues samples were normalised to the intensity of the amide II band in the second derivative spectra (intensity at 1547 cm^{-1} minus intensity at 1527 cm^{-1}). Urine samples were normalised to the area of the urea band at 1468 cm^{-1} in the absorbance spectra (integrated area between the spectrum and a linear baseline between 1487 and 1436 cm^{-1}).

2.4.2 Ramping of ATR-FTIR spectra

In general, the thickness of the samples was sufficient to fill the entire active volume above the prism surface. Hence, in order to adjust the ATR spectra for the distortion caused by the frequency-dependency of the pathlength [3], ATR data of urine samples between 1510 and 1410 cm^{-1} were ramped before fitting of components by multiplying absorbance values by a factor of (wavenumber$_{max}$/wavenumber) to allow for the inverse relation between effective pathlength and wavenumber. However, because the wavenumber range used for fitting is very narrow (100 cm^{-1}), this correction is extremely small and, in practice, does not significantly alter the fitting parameters or the relative areas of the best fits.

2.4.3 Curve fitting

Decomposition of spectra using curve fitting analysis was performed using the 'Peak Analyzer' function in Origin version 8.6. Spectra were simulated with combinations of Gaussian, Lorentzian and/or pseudo-Voigt functions. The functions used to fit the Gaussian (Eq. 2.1), Lorentzian (Eq. 2.2) and pseudo-Voigt (Eq. 2.3) components were as follows:

$$y = y_0 + \frac{A}{\sqrt[w]{x/2}} e^{-2\frac{(x-x_c)^2}{w^2}} \tag{2.1}$$

$$y = y_0 + \frac{2A}{\pi} \frac{w}{4(x-x_c)^2 + w^2} \tag{2.2}$$

$$y = y_0 + A\left[m_u \frac{2}{\pi} \frac{w}{4(x-x_c)^2 + w^2} + (1-m_u)\frac{\sqrt{4ln2}}{\sqrt{\pi w}} e^{\frac{4ln2}{w^2}(x-x_c)^2}\right] \tag{2.3}$$

where y_0 = baseline offset, A = amplitude, m_u = profile shape factor, w = full width at half maximum peak height (FWHM) and x_c = peak centre. The pseudo-Voigt function is the weighted sum of a Gaussian and a Lorentzian functions, where the profile shape factor (m_u) specifies the fraction contribution of the Lorentzian function to the overall fit. A single width (w) was used to describe Gaussian and Lorentzian components of the pseudo-Voigt function [7, 8, 97, 98].

2.4.4 Integration analysis

A simple integration analysis was used to assess the differences between spectra from different groups of samples. Once bands of potential significance were identified in the spectra, limits of integration for each band were defined. Generally, these limits were selected in order to include at least two-thirds of the band. The area of the band was then found by integration between the spectrum and a linear baseline between the defined limits according to the trapezium rule using an in-house script written in MATLAB R2012b software.

Where univariate analyses were performed, the Shapiro-Wilk test of normality was first performed on the data in OriginPro 8.6 with a significance level of 0.05. This confirmed whether or not the data were normally distributed and, hence, whether parametric or non-parametric analyses should be applied.

Mann Whitney U analysis was performed with OriginPro 8.6 software to quantitatively compare the feature areas. Two-tailed tests were applied in which the null hypothesis in each case was that there was no difference between the distributions of feature areas of the two groups in question. The alternative hypothesis was that there was a difference between the distributions. The significance level used was 0.05 which equates to 0.025 at either end, but p-values are also included to show which features are also different at a significance level of 0.01.

2.4.5 Bland-Altman analysis

Bland-Altman plots [99] were used to compare the FTIR and Jaffe methods for quantitating creatinine and the FTIR and ion-exchange chromatography methods for quantitating cystine. The mean bias for each plot was calculated and plotted as a dashed line and 95% confidence limits (mean bias +/- 1.96 standard deviations) are indicated by dotted lines.

2.4.6 Partial least squares (PLS) logistic regression

Tissue spectra were analysed by PLS logistic regression. PLS regression was applied to generate latent variables (LVs) from the original dataset (see introduction for details). In this case, the x-variables represented the absorbance values at each wavenumber, and the y-variables represented the normal pancreas or tumour class assignments. The LVs represent the variations between spectra. Each sample spectrum is compared to each LV/loading and scores are assigned based on how much each loading contributes to the given sample spectrum. A classification model was then built by performing logistic regression using the PLS LVs as covariates.

PLS logistic regression was performed using the PLS_toolbox (from Eigenvector) in MATLAB R2012b. Spectra were cut between 1800 and 950 cm^{-1} and all data (x and y) were mean centred. The default SIMPLS algorithm in the PLS_toolbox was applied for PLS regression. Leave-one-patient-out cross validation was used to assess the predictive ability of the model (part of this process utilised a script written by Liberty Foreman). The spectral region and number of LVs used to build the PLS logistic regression models were optimised on a case by case basis as detailed in results Chapter 7.

2.4.7 Calculation of sensitivities and specificities

The performances of the PLS logistic regression models were evaluated by calculation of the sensitivities, specificities and overall accuracies. Sensitivity describes the extent to which positive cases for a given group are correctly identified as such and false negatives are minimised. Sensitivity is calculated as shown in equation 2.4. Specificity describes the extent to which false positives are minimised as shown in equation 2.5. The calculation for accuracy is shown in equation 2.6.

$$Sensitivity = \frac{TP}{TP + FN} \tag{2.4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2.5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.6}$$

where TP = true positive, FP = false positive, TN = true negative and FN = false negative [46].

2.5 Electronic structure modelling of urea-water and urea-urea structures

Infrared spectra of urea-water and urea-urea molecular ensembles were modelled. Firstly, nominal molecular models were built by stepwise addition of water or urea using Facio 18.6.2 [100]. Structures were then energy minimised using Gaussian 09 software [101] with the B3LYP density function in conjunction with the 6-31G(d) basis set, using the UCL Legion supercomputer facility. Calculations were performed in vacuum. Vibrational spectra were then simulated with the energy-minimised structures. Simulated frequencies were visualized with Facio 18.6.2 after scaling [102] with the Facio default factor of 0.9614. The predicted infrared spectrum of creatinine was modelled in the same manner.

2.6 Urine analyses

2.6.1 Generation of urea and creatinine calibration curves

Solutions of urea (25 mM – 200 mM) and of creatinine (2 mM – 30 mM) were prepared in double distilled water. A mixture of 100 mM urea and 10 mM creatinine was also prepared. 20 µL aliquots were placed on the ATR prism and the FTIR spectra were recorded as described. Signal intensities were used to generate calibration curves for these compounds to enable quantitation of urea and creatinine in urine.

2.6.2 Generation of a cystine calibration curve

A 32 mM stock solution of cystine was prepared by dissolving microcrystalline cystine in 1M HCl; 50 mM sodium phosphate was added followed by 1M NaOH to neutralise, generating a very fine precipitate of cystine. The suspension was adjusted to

pH 7 with NaOH. This stock suspension was mixed by vortexing and diluted with urine from healthy human donors to give 1 mL samples of 0 to 3 mM total cystine. After mixing the insoluble fractions were separated and ATR-FTIR spectra of dried aliquots were recorded as described. A calibration curve was generated based on signal intensities to enable quantitation of cystine in urine (intensity 1296 cm^{-1} minus 1280 cm^{-1}).

2.6.3 Quantitation of creatinine by the Jaffe reaction

The Jaffe reaction is the current method used in standard clinical practice for quantitating urinary creatinine. The Jaffe working reagent was prepared by mixing five parts of 200 mM NaOH with one part of 40 mM picric acid. Absorbance at 510 nm was monitored using a Shimadzu 50000 UV/Visible spectrophotometer at 25 °C. 0-30 mM creatinine standards and urine samples were diluted 50-fold. Three millilitres of working reagent were placed in a glass cuvette and the rate of change of absorbance at 510 nm was recorded on addition of 200 µL diluted creatinine or urine. The initial rates of reaction after addition of known quantities of creatinine were plotted versus creatinine concentrations to generate a calibration curve for estimation of urine creatinine levels.

2.6.4 Standard clinical quantitation of cystine

Ion-exchange chromatography is the gold standard in clinical amino acid analysis [103, 104]. Urine samples were sent to the Camelia Botnar Laboratories at Great Ormond Street Hospital for urinary cystine analysis using a Biochrom 30+ Amino Acid Analyzer. Samples were warmed to room temperature, deproteinised with sulphosalycic acid containing norleucine internal standard, then loaded onto a column linked to sulphonate groups to permit charge-charge interaction. Amino acids were sequentially eluted over 2-3 hours with increases in temperature, pH and ion strength using lithium citrate.

2.6.5 Vapour phase measurements of trimethylamine (TMA)

Ten millimolar and 1 mM solutions of TMA in double distilled water were prepared. One hundred microlitre aliquots were placed inside the vapour-measurement device and spectra recorded as described.

2.6.6 Solution phase measurements of TMA and TMA N-oxide

Solutions of 50 mM TMA in 50 mM phosphate buffer pH 7.25; 25 mM TMA N-oxide in 25 mM phosphate buffer pH 7; and 5 mM TMA N-oxide in 5 mM phosphate buffer pH 7 were prepared. Spectra were recorded with a Bruker IFS/66S FTIR spectrometer as described except that 1000 interferograms were averaged. The absorbance spectra were computed versus a background of phosphate buffer (at pH and concentration corresponding to each sample) or of the clean prism surface (5000 interferograms).

Chapter 3

Model compounds

3.1 Introduction

Since most biological compounds are infrared active, spectra of biological samples tend to be complex, with many overlapping absorbance bands. This can make deconvolution of the spectra difficult and it is not always possible to make assignments based on only a single band. In some cases, it can be possible to positively identify a specific compound if several associated absorbance bands are present at the expected relative intensities. Model spectra of individual compounds under different conditions (i.e. in solution *versus* dried or at defined pH values) can be compared to complex sample spectra to determine whether the unique spectral band patterns can be seen.

When comparing sample spectra to model compound spectra it is important that the environment of the model matches the environment of the sample as closely as possible. For example, hydrogen bonding can cause an upshift in absorbance band positions, therefore spectra of compounds in solution can differ from spectra of the same compound dried. Furthermore, interactions with surrounding molecules can influence the relaxation rate and/or coherence of vibrating molecules and lead to differences in lineshapes between solid, liquid and gaseous states. An additional consideration is whether the compound of interest contains any ionisable groups, in which case it is important to consider the pH/ionic strength of the sample and model, as different protonation states will have different IR spectra.

As well as aiding in the identification of specific compounds of interest, model spectra of dominant compounds in a sample can be subtracted from sample spectra to better reveal underlying minor components that would otherwise be hidden. Model spectra can also be used to generate calibration curves, based on intensity or area of absorbance bands, that can be used to quantitate specific compounds in some types of biological samples.

3.2 Urine-related compounds

After water, urea is the next most dominant component in urine. Creatinine is also a major component and is typically used clinically for normalising urine concentrations. Other constituents include chlorides, ammonium ions, phosphates, sulphates and other minor metabolites including organic compounds and organic and inorganic salts [105–107]. Some diseases can result in elevated levels of certain metabolites, for example, patients with kidney stone disease may have raised urinary oxalate, urate, uric acid, carbonates or cystine, and diabetic patients may have excess glucose in their urine.

Breakdown products of certain foods and drugs may also be excreted in urine at detectable levels [52, 108].

For detecting and quantitating major components, such as urea and creatinine, wet spectra of urine are adequate. In some cases it is necessary to dry the urine to remove the dominant absorbance bands of water and increase the relative concentration of more minor metabolites in the active area of the prism to increase their absorbance band intensities. Therefore, a solution spectrum and a dried spectrum were recorded for each compound. In some cases a spectrum of the solid, crystalline compound was also recorded. Compounds with ionisable groups were buffered with potassium phosphate to pH 6 or 7, as this is the approximate pH range of a typical urine spectrum. In such cases, pH-dependent spectra of the phosphate buffer can be subtracted to leave the pure model compound spectrum.

Figures 3.1 to 3.5 show the model spectra of the main urinary constituents, including urea, creatinine, ammonium chloride (adjusted to pH 6 with phosphate buffer), potassium phosphate (pH6) and potassium sulphate (pH 7). For each compound a solution spectrum after subtraction of pure water and water vapour spectra, and the corresponding dried spectrum, are shown, along with the second derivative transformations of each absorbance spectrum (shown immediately below).

Figure 3.1 shows model spectra of solution and dried urea. Urea is a waste product formed during the metabolism of nitrogen-containing compounds, such as dietary protein, in the body [106]. The IR spectrum of urea is relatively simple in the 'fingerprint' region. Distinct bands can be seen at 1468 cm^{-1} (solution) or 1464 cm^{-1} (dried) due to the antisymmetric CN stretch, and at 1159/1152 cm^{-1} due to NH rocking vibrations. Three overlapping bands are also seen between 1700 and 1500 cm^{-1} in absorbance spectra (symmetric and antisymmetric NH bend and CO stretch) [109, 110], which become reasonably well resolved in the second derivative transformations.

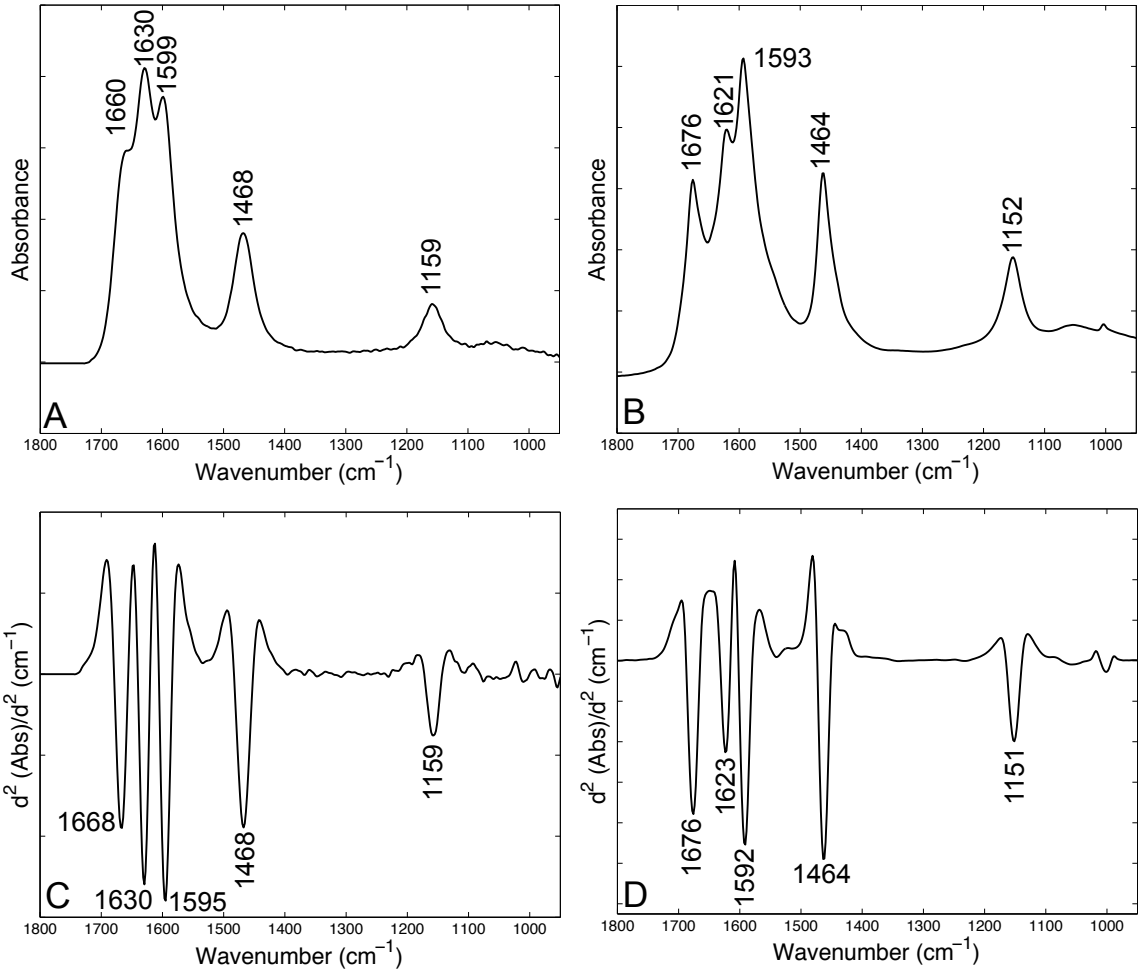


Figure 3.1: Model spectra of urea. *(A) Solution absorbance spectrum after subtraction of pure water and water vapour spectra. (B) Dried absorbance spectrum. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

Creatinine is a larger molecule than urea, and is produced from the breakdown of creatine phosphate in muscle [106]. It has a more complex absorbance spectrum (Figure 3.2) with prominent bands at 1492, 1348 and 1253 cm^{-1} in the solution absorbance spectrum, as well as other overlapping band between 1750 and 1550 cm^{-1} and lower intensity bands at 1427 and 1099 cm^{-1} (Figure 3.2 A).

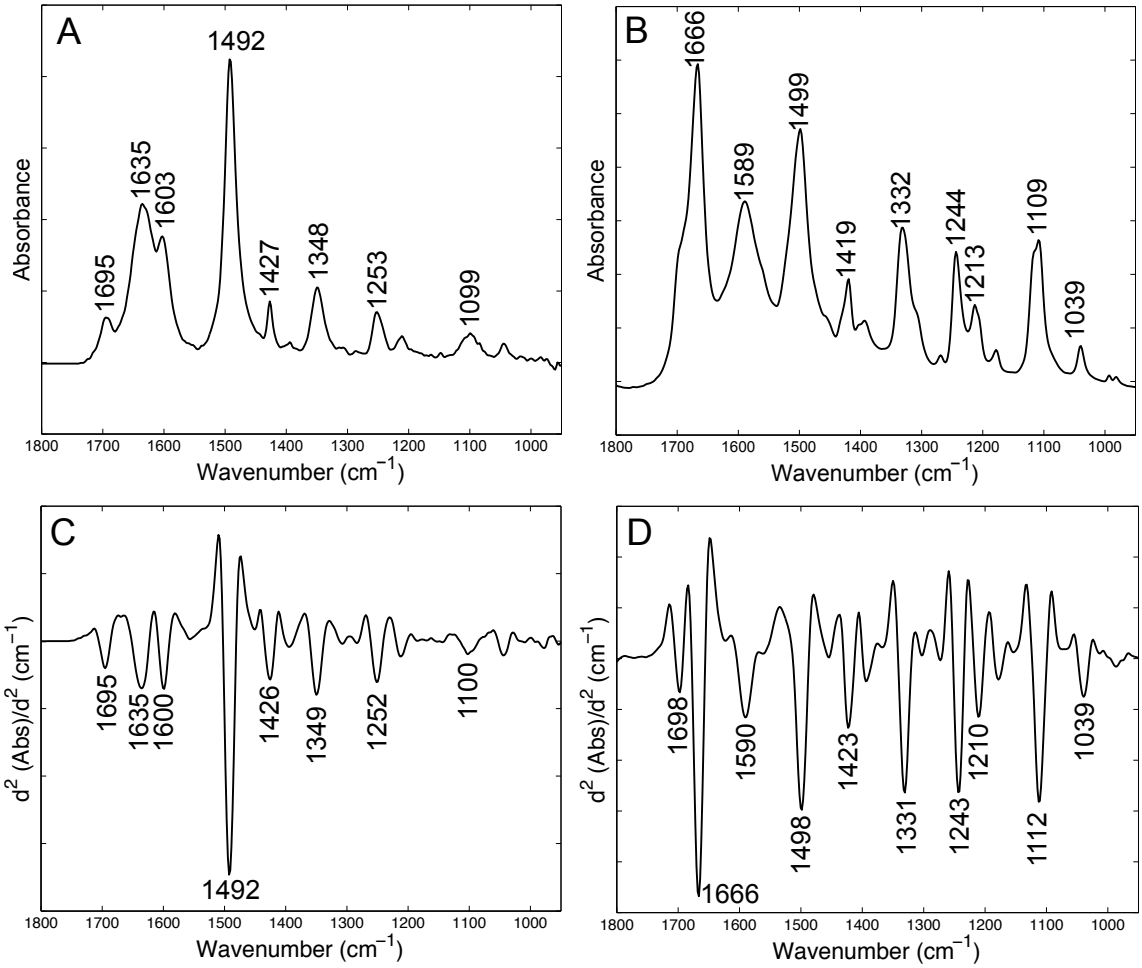


Figure 3.2: Model spectra of creatinine. *(A) Solution spectrum after subtraction of pure water and water vapour spectra. (B) Dried spectrum. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

Ammonium ions (Figure 3.3) are produced during renal metabolism [111]. Its main absorbance is at around 1406 cm^{-1} with a shoulder at about 1441 cm^{-1}. The absorbance bands between 1150 and 950 cm^{-1} are mainly attributed to the phosphate buffer, which has not been subtracted in this case since the absorbance bands of phosphate do not overlap with those of ammonium ions.

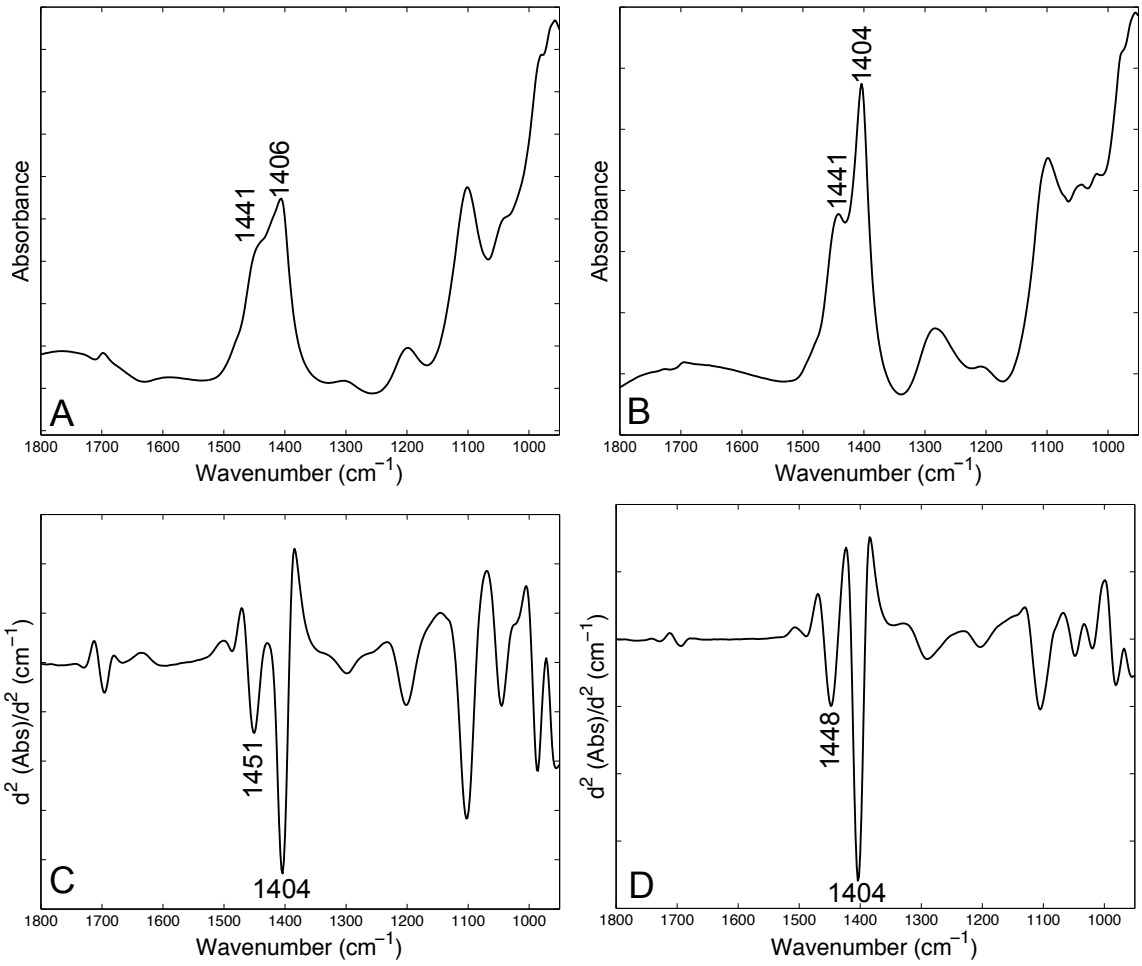


Figure 3.3: Model spectra of ammonium ions. *All solutions were buffered to pH6 with phosphate buffer. (A) Solution spectrum of ammonium chloride after subtraction of pure water and water vapour spectra. (B) Dried spectrum. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

Phosphate (Figure 3.4) and sulphate (Figure 3.5) ions absorb strongly between 1200 and 900 cm^{-1}, with the rest of the fingerprint regions up to 1800 cm^{-1} being relatively featureless.

Phosphate ions in solution have strong absorbance bands at 1157 and 1076 cm^{-1} (Figure 3.4 A) due to the PO stretch [4]. The positions of these bands are very sensitive to changes in pH and can be used to deduce the pH of a sample.

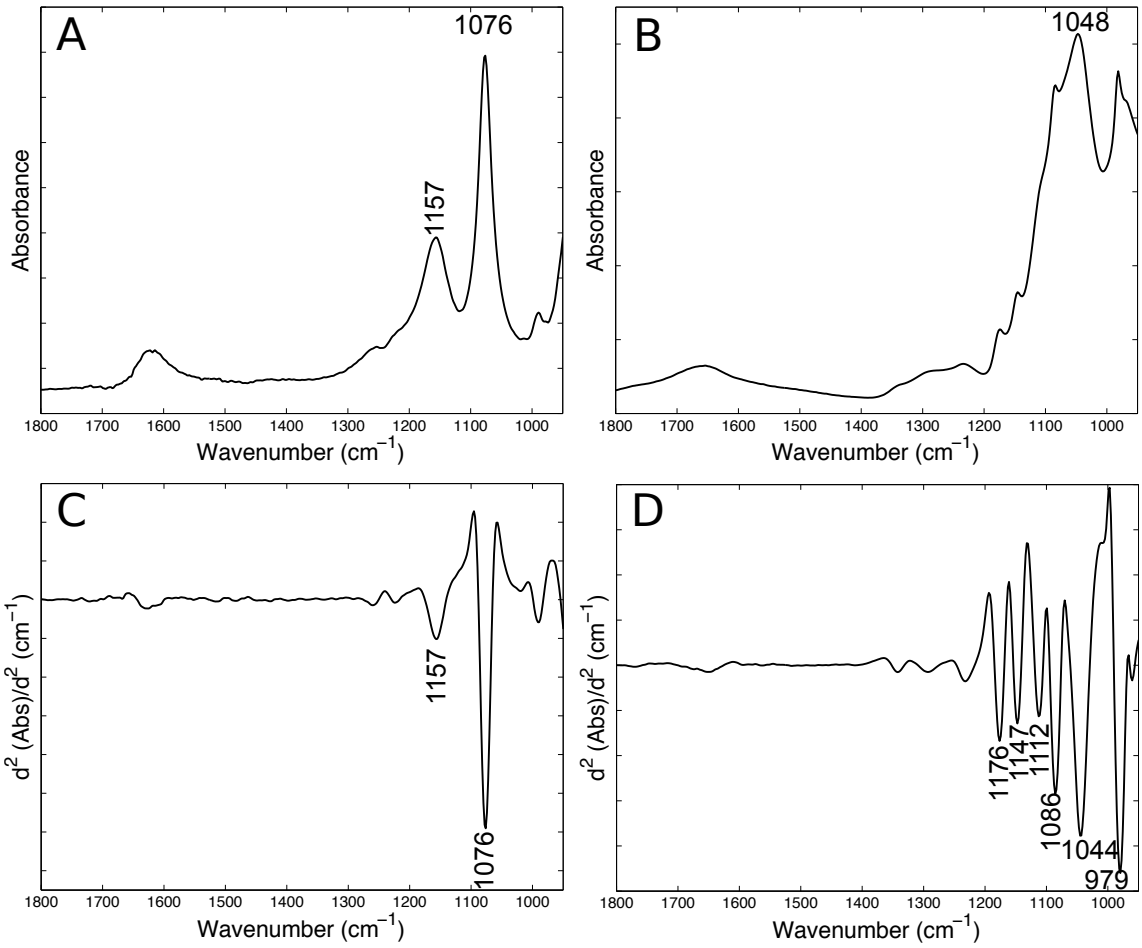


Figure 3.4: Model spectra of potassium phosphate at pH 6. *(A) Solution spectrum after subtraction of pure water and water vapour spectra. (B) Dried spectrum. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B. The solution was adjusted to pH6 with potassium hydroxide.*

Sulphate has a strong broad band at 1085 cm^{-1} in the solution absorbance spectrum (Figure 3.5 A).

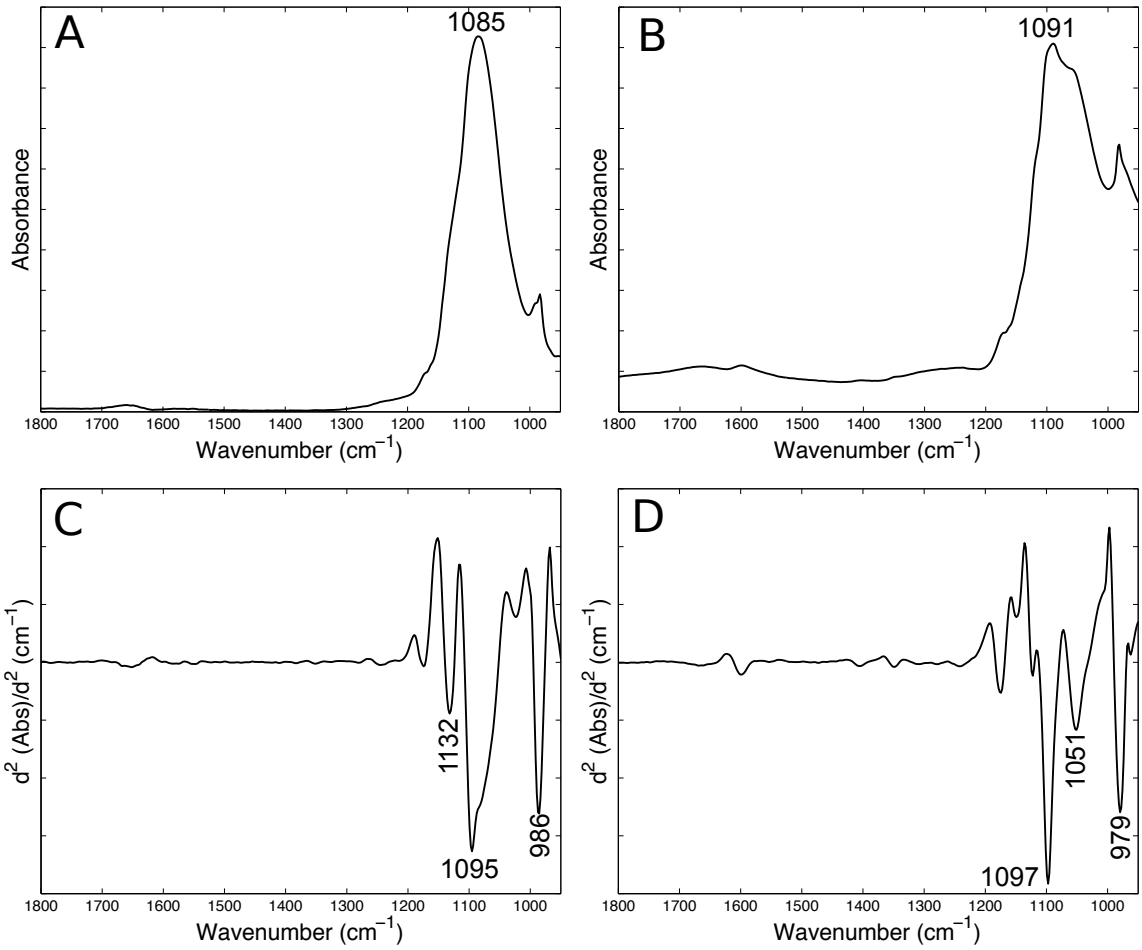


Figure 3.5: Model spectra of potassium sulphate at pH 7. *(A) Solution spectrum after subtraction of pure water and water vapour spectra. (B) Dried spectrum. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B. The solution was adjusted to pH7 with potassium hydroxide.*

In addition to these major urinary constituents there are a number of more minor metabolites and disease-associated compounds that are found in some urine samples. These include compounds such as sugars, stone-forming compounds and other indicators of kidney malfunction, such as protein [108] (model protein spectra are shown in the next section describing tissue-related compounds).

Figure 3.6 shows the model spectra of glucose in solution after subtraction of pure water and water vapour spectra, and dried. The second derivative transformations of the absorbance spectra are also shown immediately below. The main absorbance bands are between 1200 and 950 cm^{-1} attributed to CO stretching and bending vibrations [4].

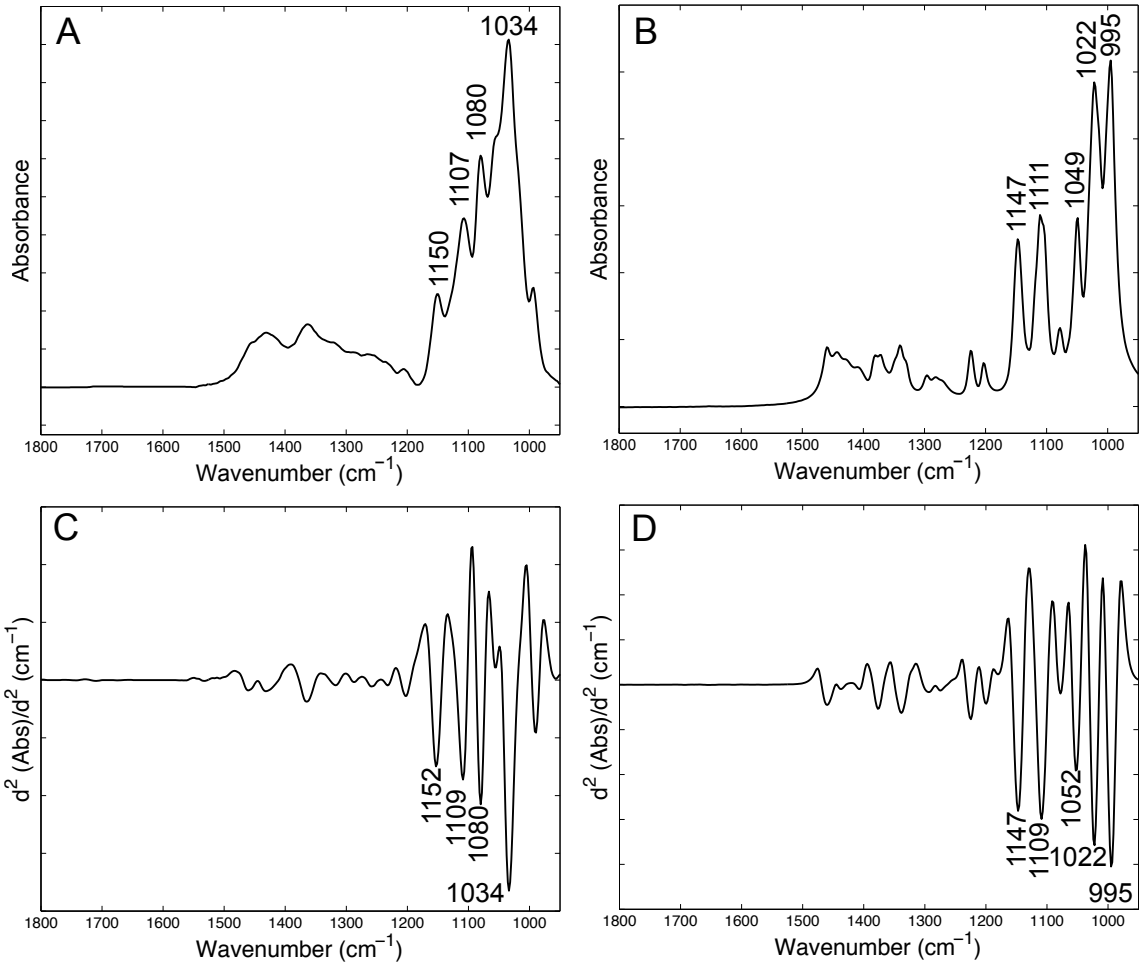
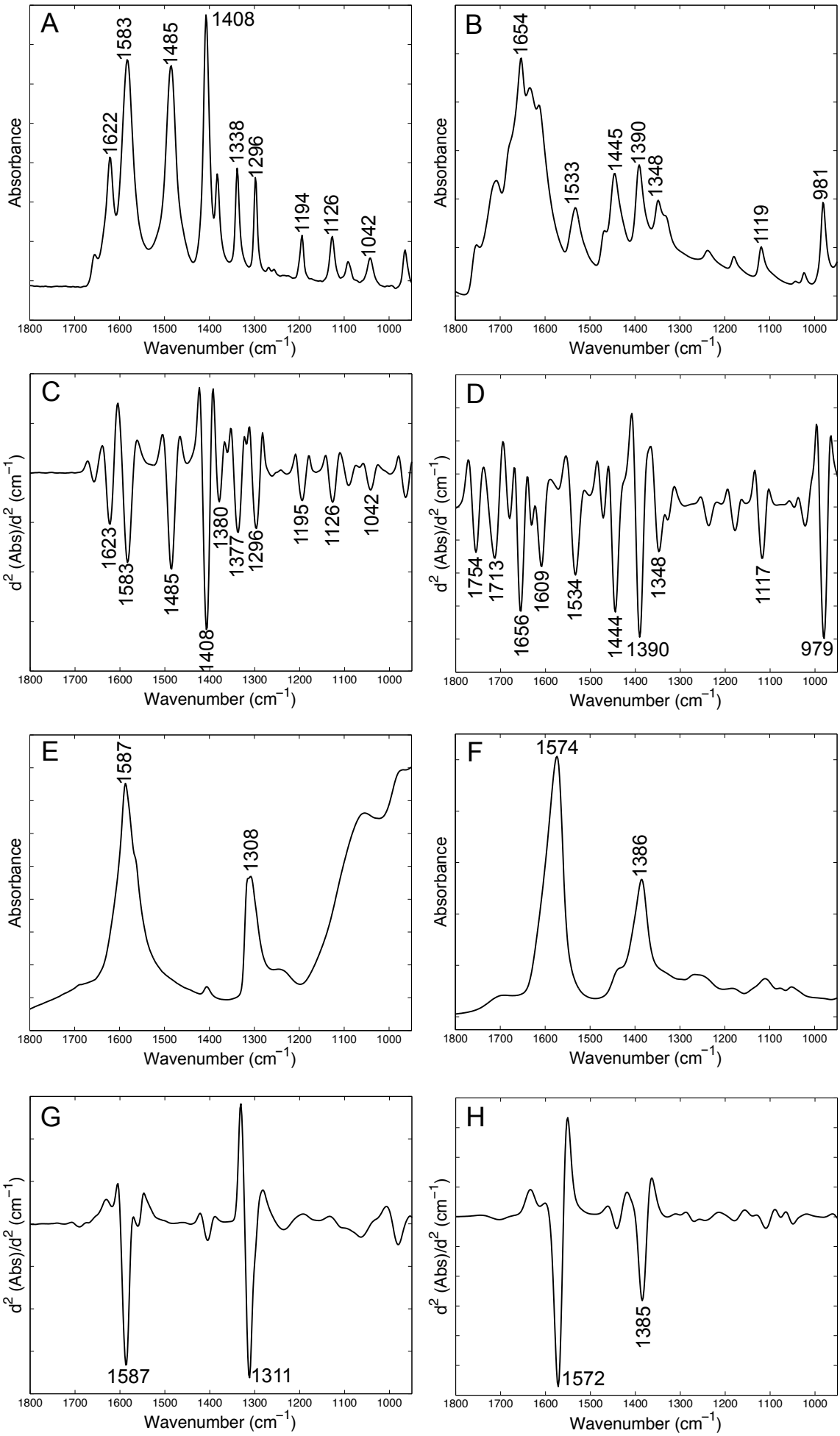


Figure 3.6: Model spectra of glucose. *(A) Solution spectrum after subtraction of pure water and water vapour spectra. (B) Dried spectrum. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

Figure 3.7 shows the model spectra of selected stone-forming compounds. Since these are particularly hydrophobic in nature their absorption band peak positions do not shift significantly between wet and dried states so only dried or solid spectra of each compound have been shown.


A
1622
1583
1485
1408
1338
1296
1194
1126
1042
Absorbance
Wavenumber (cm^{-1})
B
1654
1533
1445
1390
1348
1119
981
Absorbance
Wavenumber (cm^{-1})
C
1623
1583
1485
1408
1380
1377
1296
1195
1126
1042
d^2 (Abs)/d^2 (cm^{-1})
Wavenumber (cm^{-1})
D
1754
1713
1656
1609
1534
1444
1390
1348
1117
979
d^2 (Abs)/d^2 (cm^{-1})
Wavenumber (cm^{-1})
E
1587
1308
Absorbance
Wavenumber (cm^{-1})
F
1574
1386
Absorbance
Wavenumber (cm^{-1})
G
1587
1311
d^2 (Abs)/d^2 (cm^{-1})
Wavenumber (cm^{-1})
H
1572
1385
d^2 (Abs)/d^2 (cm^{-1})
Wavenumber (cm^{-1})

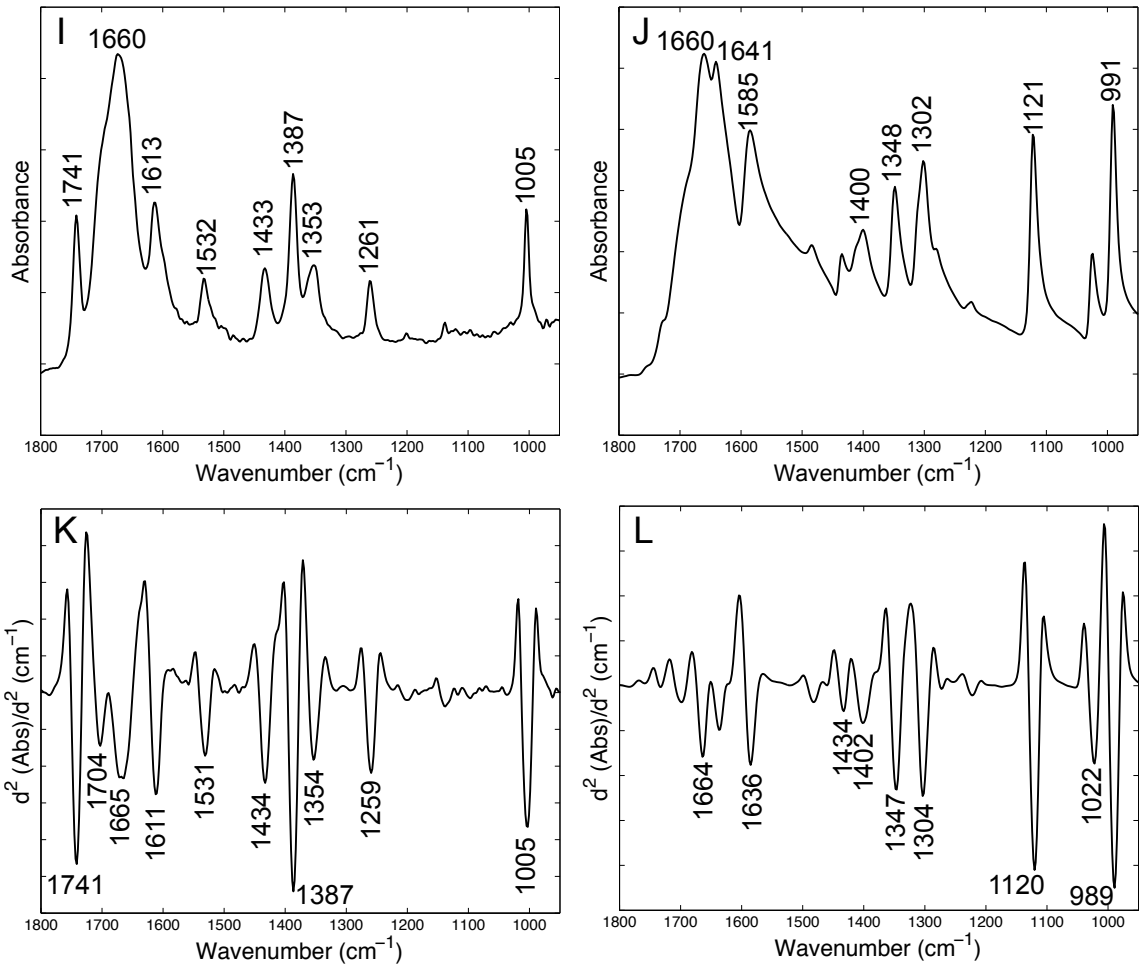


Figure 3.7: Model spectra of selected stone-forming compounds. *(A) Dried cystine. (B) Solid dihydroxyadenine. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B. (E) Dried solution of oxalate, buffered to pH 6 with phosphate buffer. (F) Dried solution of citrate, buffered to pH 6 with phosphate buffer. (G) Second derivative of spectrum in E. (H) Second derivative of spectrum in F. (I) Solid sodium urate. (J) Solid uric acid. (K) Second derivative of spectrum in I. (L) Second derivative of spectrum in J.*

3.3 Tissue-related compounds

Tissue spectra are more complex than urine spectra, since they contain many different types of proteins, carbohydrates, lipids, small metabolites, DNA and RNA. Spectra are usually dominated by the amide I and amide II absorption bands of protein, although it is usually not possible to distinguish between different types of protein, since all are made up of different combinations of the same amino acids in approximately equal distribution. The absorbance signals of the amino acid side chains are therefore averaged in the final protein spectrum. It is sometimes possible to distinguish proteins that contain a particularly high proportion of a particular amino acid, or if covalently bound

to another type of compound, such as a sugar. In fresh tissue, water also contributes significantly to the absorbance spectrum, although not to the same extent as is seen in wet urine spectra. In dried tissue, partial rehydration can occur when exposed to atmospheric moisture. Since this can have a significant effect on infrared spectra of some compounds it is important to control the atmospheric humidity during measurements.

Figure 3.8 shows the model absorbance spectrum and second derivative transformation of solid bovine serum albumin (BSA), as a general model for protein. The amide I band is centred around 1648 cm^{-1} and the amide II band is centred around 1533 cm^{-1} [4]. The amide I or amide II bands can be used to normalise tissue spectra, however since the amide I band overlaps with a strong absorption band of water normalisation to the amide II band is the preferred method.

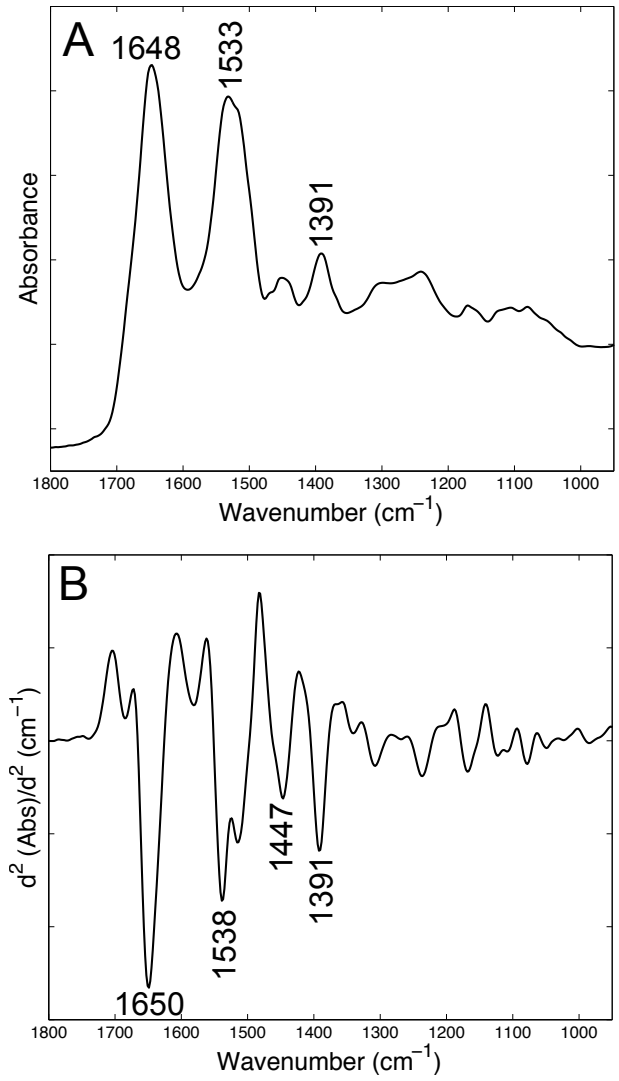


Figure 3.8: Model spectrum of solid bovine serum albumin. *(A) Absorbance spectrum. (B) Second derivative of absorbance spectrum.*

Collagen is an extracellular protein and the main constituent of connective tissue. Although it is generally not possible to distinguish different types of protein by IR spectroscopy as the overall amino acid composition is often very similar, it is potentially possible to detect differences in the collagen spectrum due to the high content of hydroxyproline (Figure 3.9). The model spectra of hydroxyproline can also be seen in Figure 3.16. Changes in collagen content of different types of breast tissue were identified by Walsh *et al.* (2012) [90].

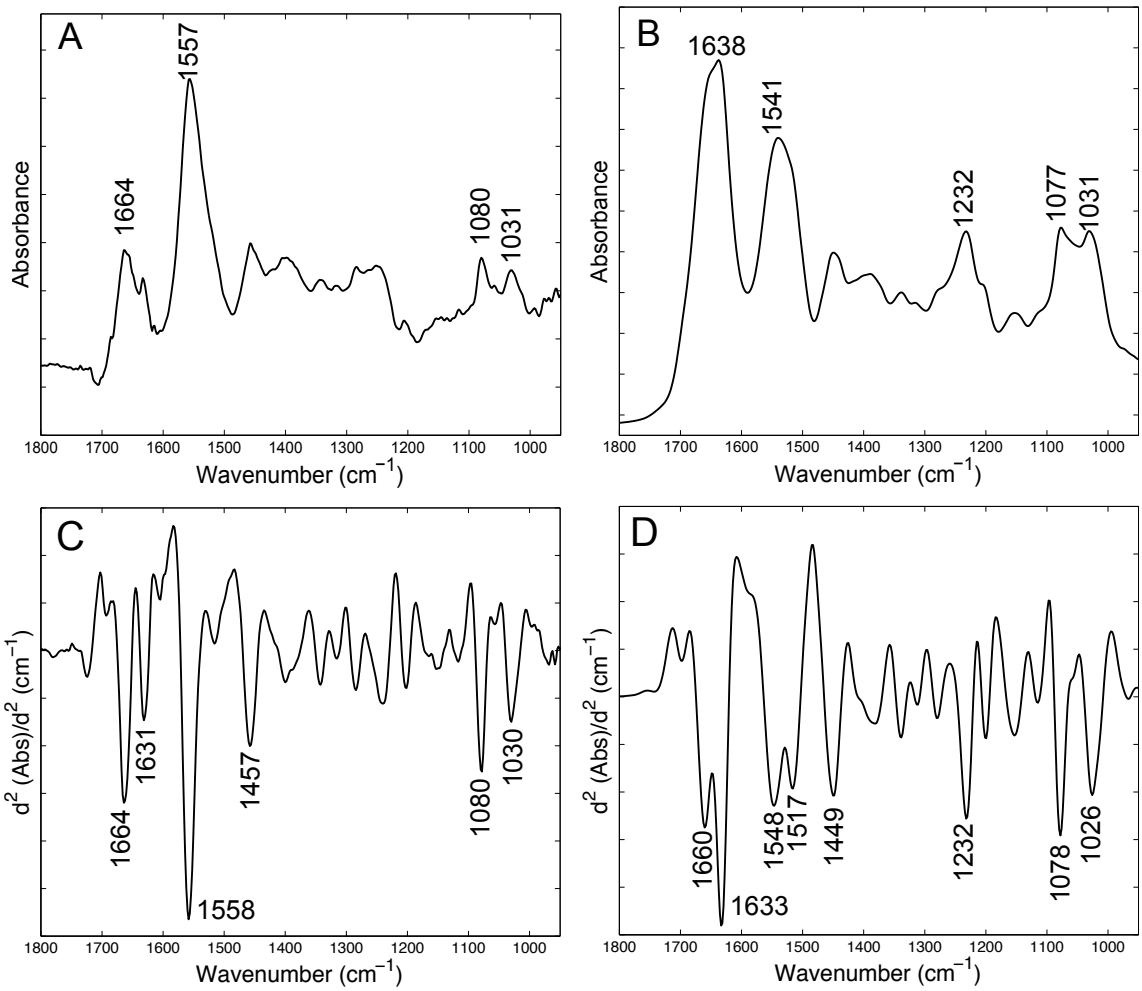


Figure 3.9: Model spectra of collagen. *(A) Solution spectrum after subtraction of pure water and water vapour spectra. (B) Dried spectrum. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

Creatine is found in high concentrations in human muscle cells, where it is used to store and liberate energy via ATP. It is also found in high concentrations in the blood as it is transported from its site of synthesis, primarily in the kidney and liver, to muscle (and brain) cells. Its model spectra are shown in Figure 3.10.

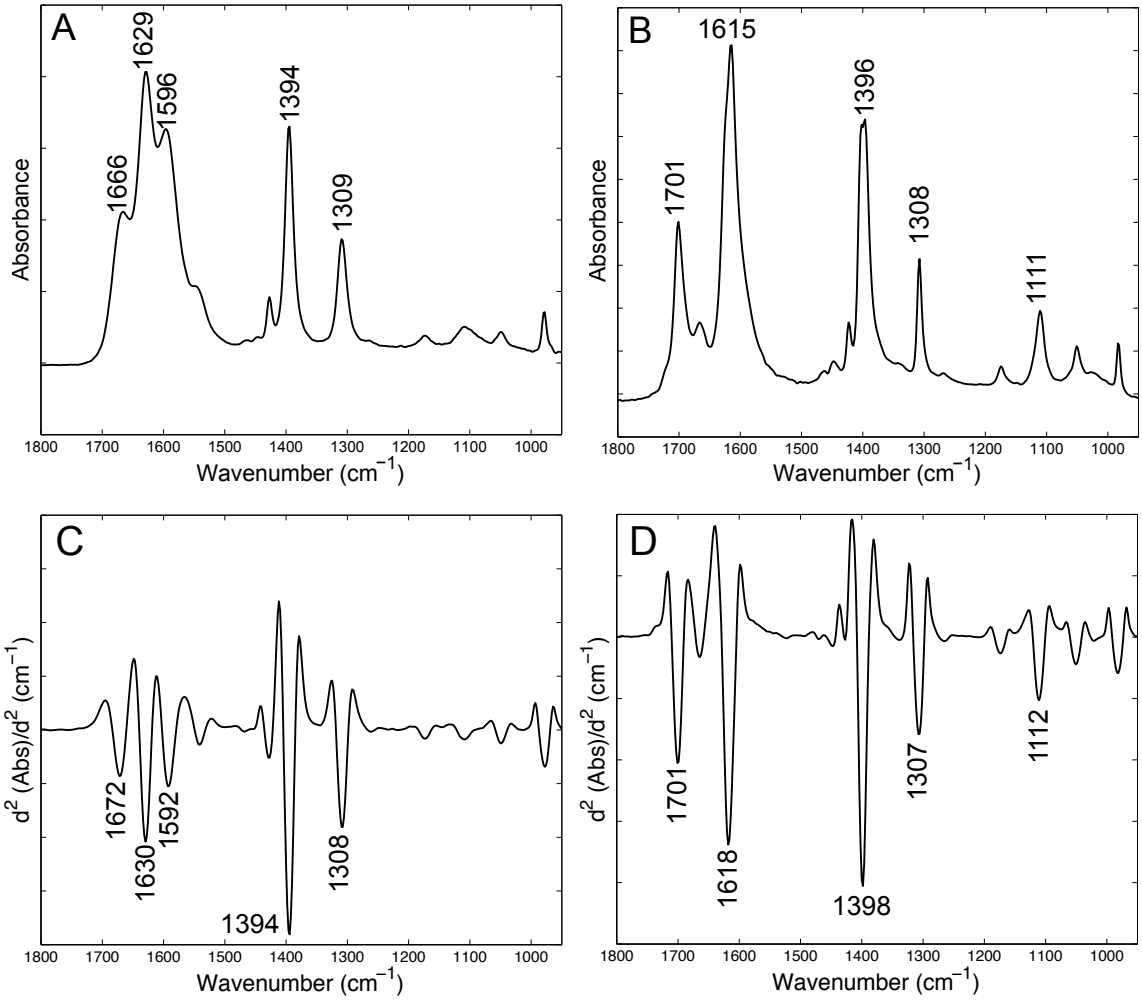


Figure 3.10: Model spectra of creatine. *(A) Solution spectrum after subtraction of pure water and water vapour spectra. (B) Dried spectrum. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

Lipids are found in cell membranes and changes in lipid content of tissue has been correlated with various cancers [20,31,35,112]. The model spectra of phosphatidylcholine and cholesterol are shown in Figures 3.11 and 3.12, respectively. As well as the fingerprint regions from 1800 to 950 cm^{-1}, spectra showing absorbances up to 4000 cm^{-1} are also included (Figures 3.11 A & C and 3.12 A & C) since lipids have strong absorbance bands between 3000 and 2750 cm^{-1}, attributed to CH stretching vibrations [4], which are regions that are relatively free of contributions from other compounds.

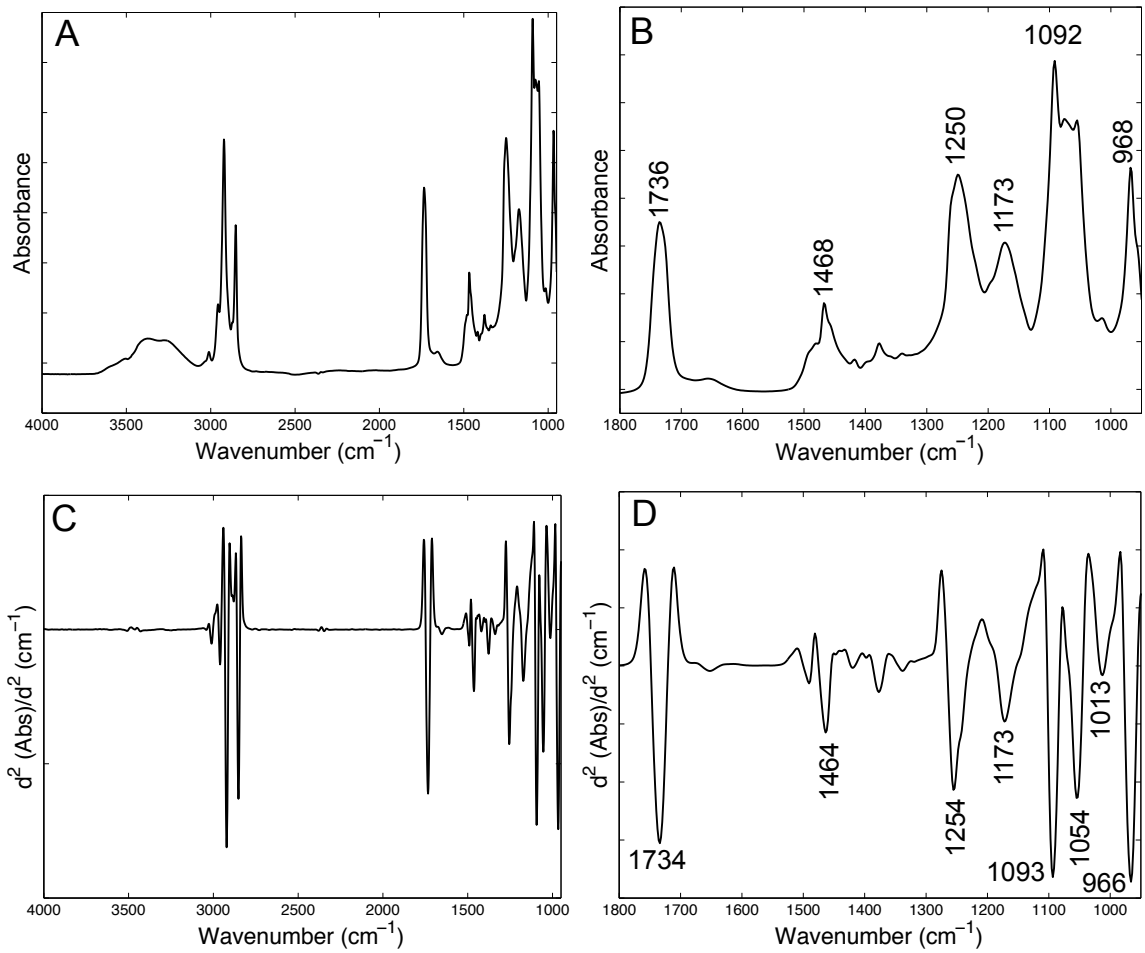


Figure 3.11: Model spectra of phosphatidylcholine. *(A) Full spectrum from 4000 - 950 cm^{-1}. (B) Fingerprint region only, 1800 - 950 cm^{-1}. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

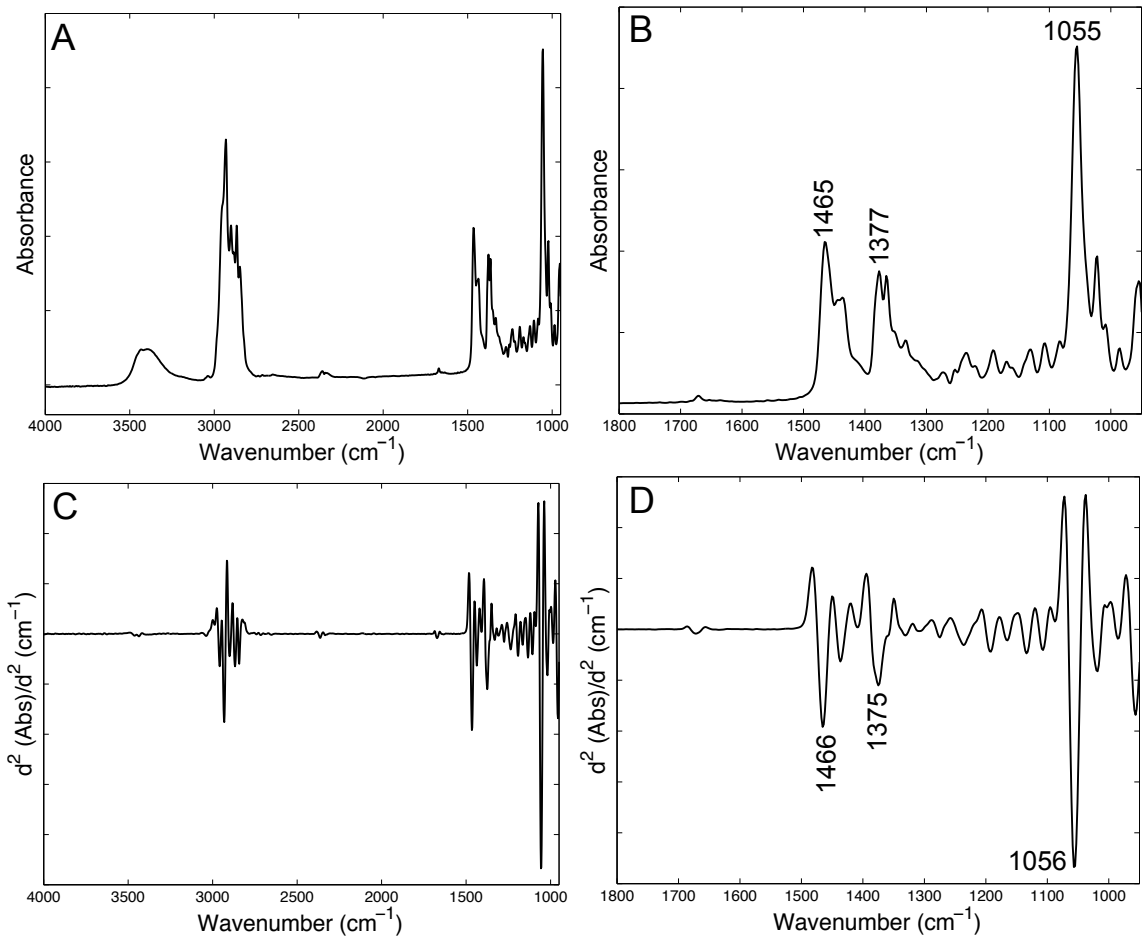


Figure 3.12: Model spectra of cholesterol. *(A) Full spectrum from 4000 - 950 cm^{-1}. (B) Fingerprint region only, 1800 - 950 cm^{-1}. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

Glycogen is a glucose storage polysaccharide found mainly in liver and muscle cells. Many studies have correlated changes in tissue glycogen content with cancer [18, 31, 35, 55, 56, 65, 66, 87]. Unsurprisingly, the absorbance spectrum of glycogen (Figure 3.13) is similar to that of glucose (Figure 3.6), with the main absorbance bands being found between 1200 and 950 cm^{-1} due to CH stretching and bending, OH bending and CC stretching vibrations [4].

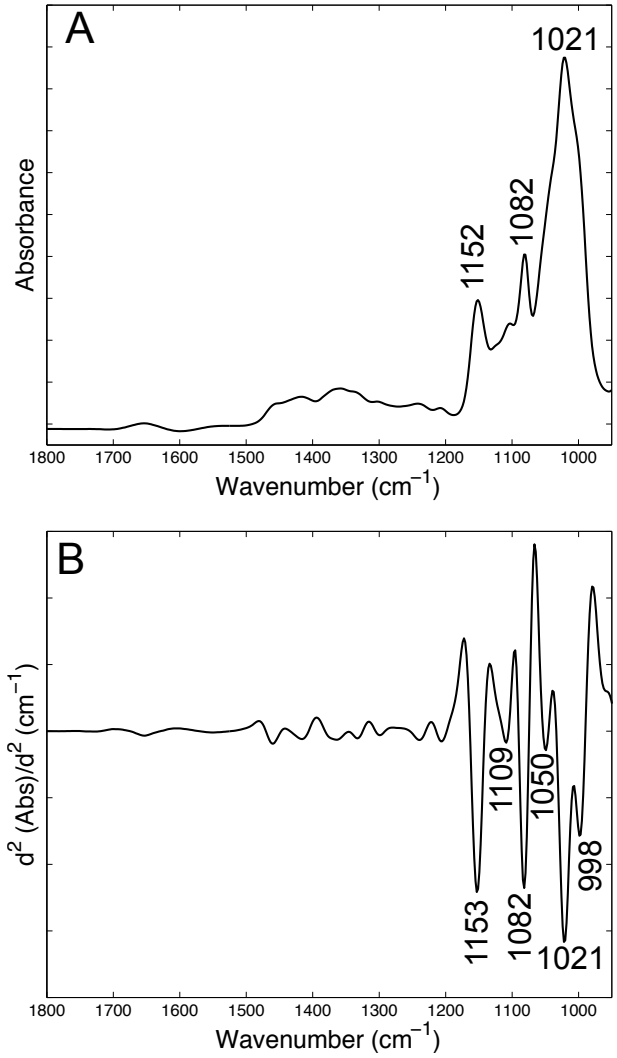


Figure 3.13: Model spectra of glycogen. *(A) Absorbance spectrum. (B) Second derivative of absorbance spectrum.*

Figure 3.14 shows the model spectra of RNA and DNA (DNA spectral data provided by Liberty Foreman). Strong absorbance bands can be seen the the 1200-950 cm^{-1} region corresponding to phosphate and are likely to be contributed to by the PO_2^- stretching of the backbone and CO stretching vibrations [4]. The broadening of all the bands in the fingerprint region suggests that some water remained in the sample during measurement. Changes in RNA/DNA of cancerous tissue have been reported by a number of research groups [31,35,86,88,113], with specific bands of DNA being identified at 970 and 1085 cm^{-1} [88], 1020 cm^{-1} [18,31,55], 1092 and 1248 cm^{-1} [35], and at 1121 cm^{-1} for RNA [18,31,55].

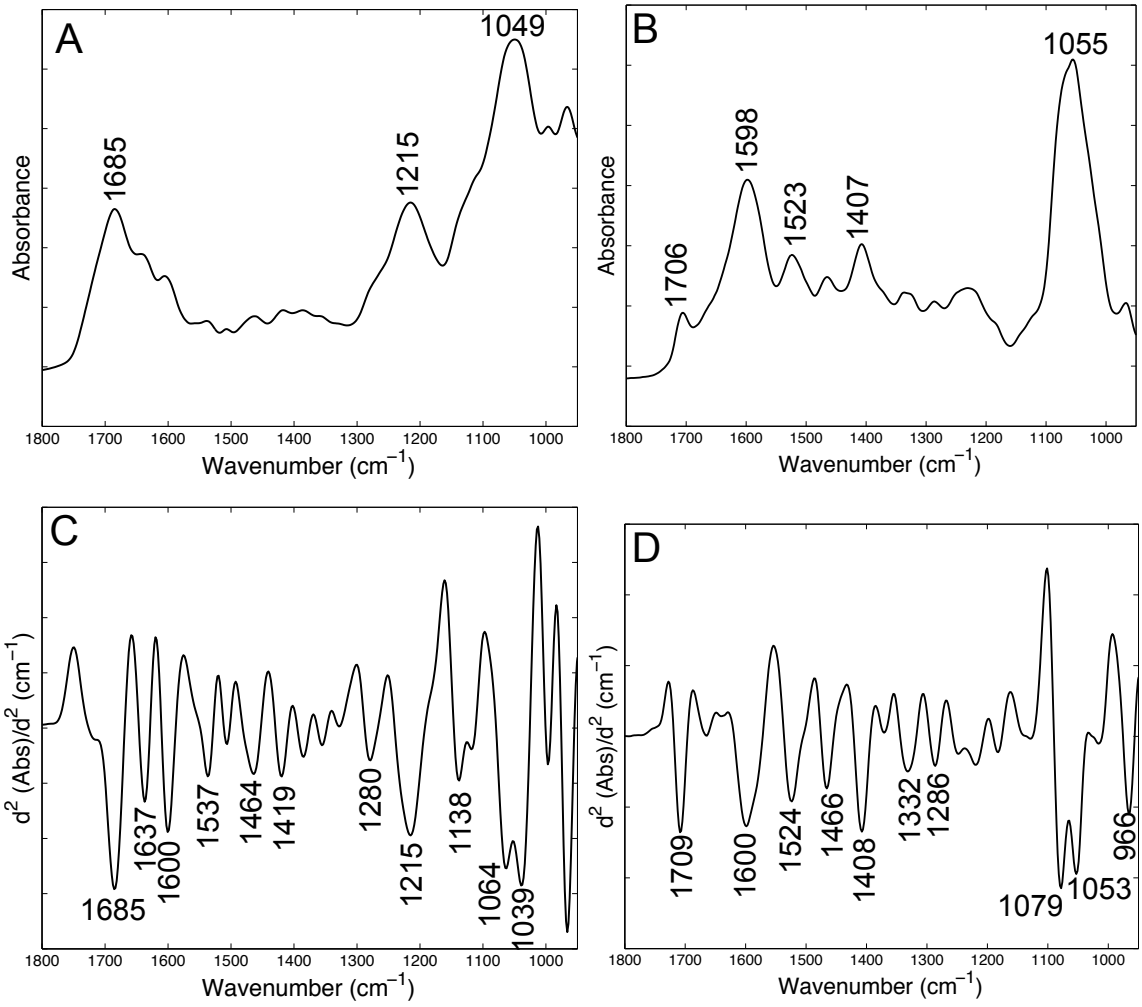


Figure 3.14: Model spectrum of RNA and DNA. *(A) Absorbance spectrum of RNA. (B) Absorbance spectrum of DNA (DNA spectral data provided by Liberty Foreman). (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

Mucins are proteins secreted by epithelial cells and are the main constituent of mucus. They are highly glycosylated, which helps distinguish their IR spectra from that of other proteins. Their spectra (Figure 3.15) contain many absorbance bands in the fingerprint region between 1800 and 950 cm^{-1}, with strong absorbance signals between 1150 and 950 cm^{-1} due to the sugars.

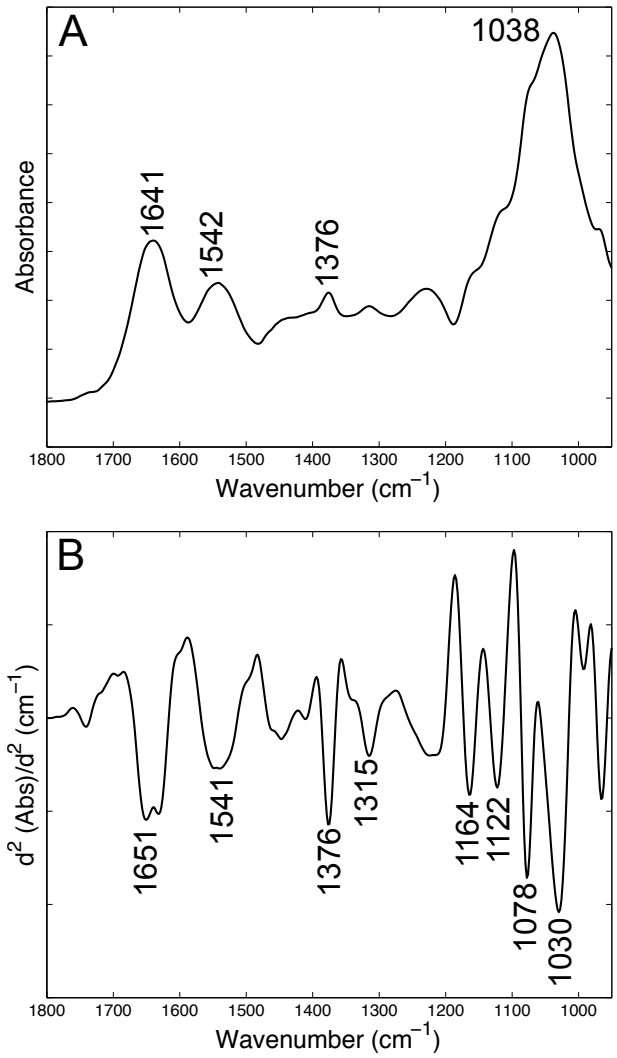


Figure 3.15: Model spectra of mucin. *(A) Absorbance spectrum. (B) Second derivative of absorbance spectrum.*

Figures 3.16 to 3.17 show the model spectra of selected amino acids with distinctive IR spectra. If a protein contains a high percentage of such amino acids it may be possible to detect their IR signals in complex mixtures.

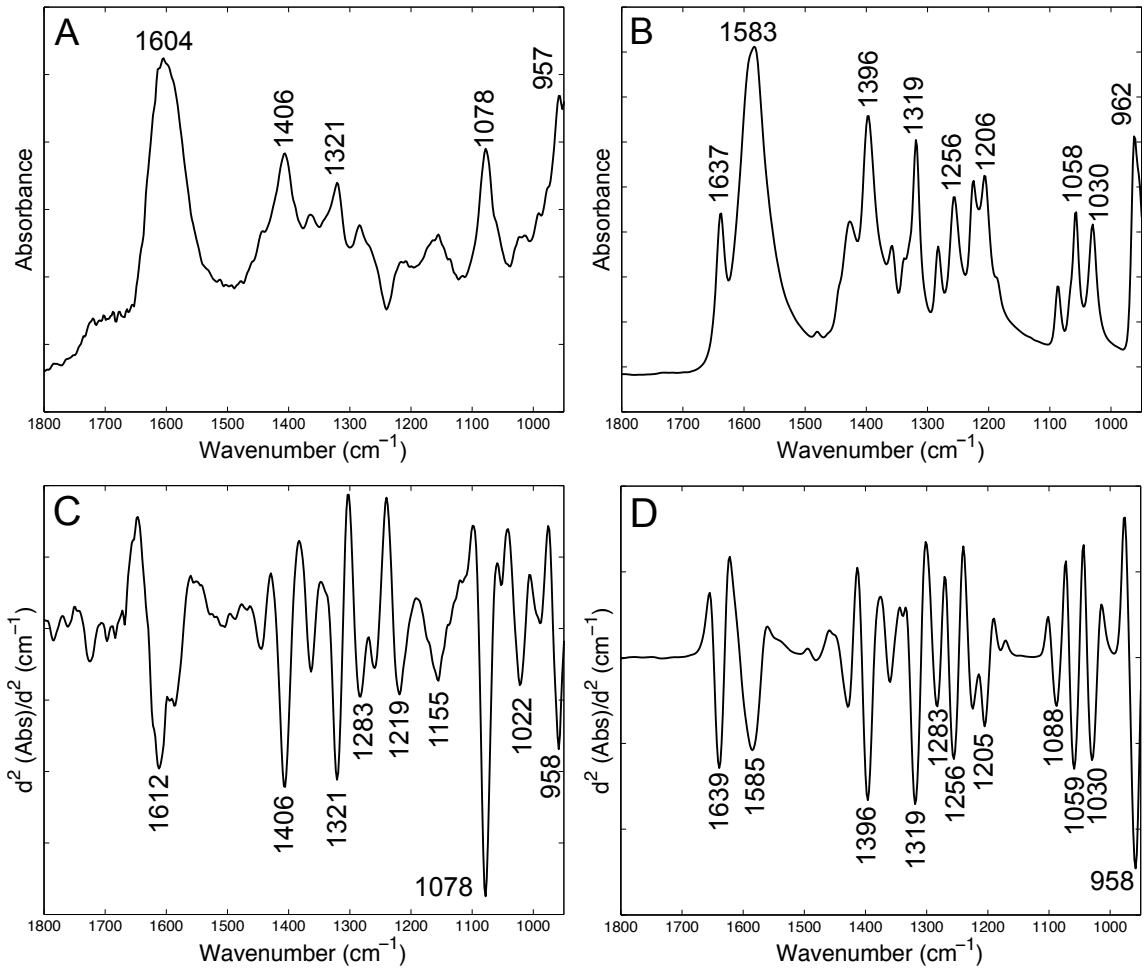


Figure 3.16: Model spectra of hydroxyproline. *(A) Solution spectrum after subtraction of pure water and water vapour spectra. (B) Solid spectrum. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

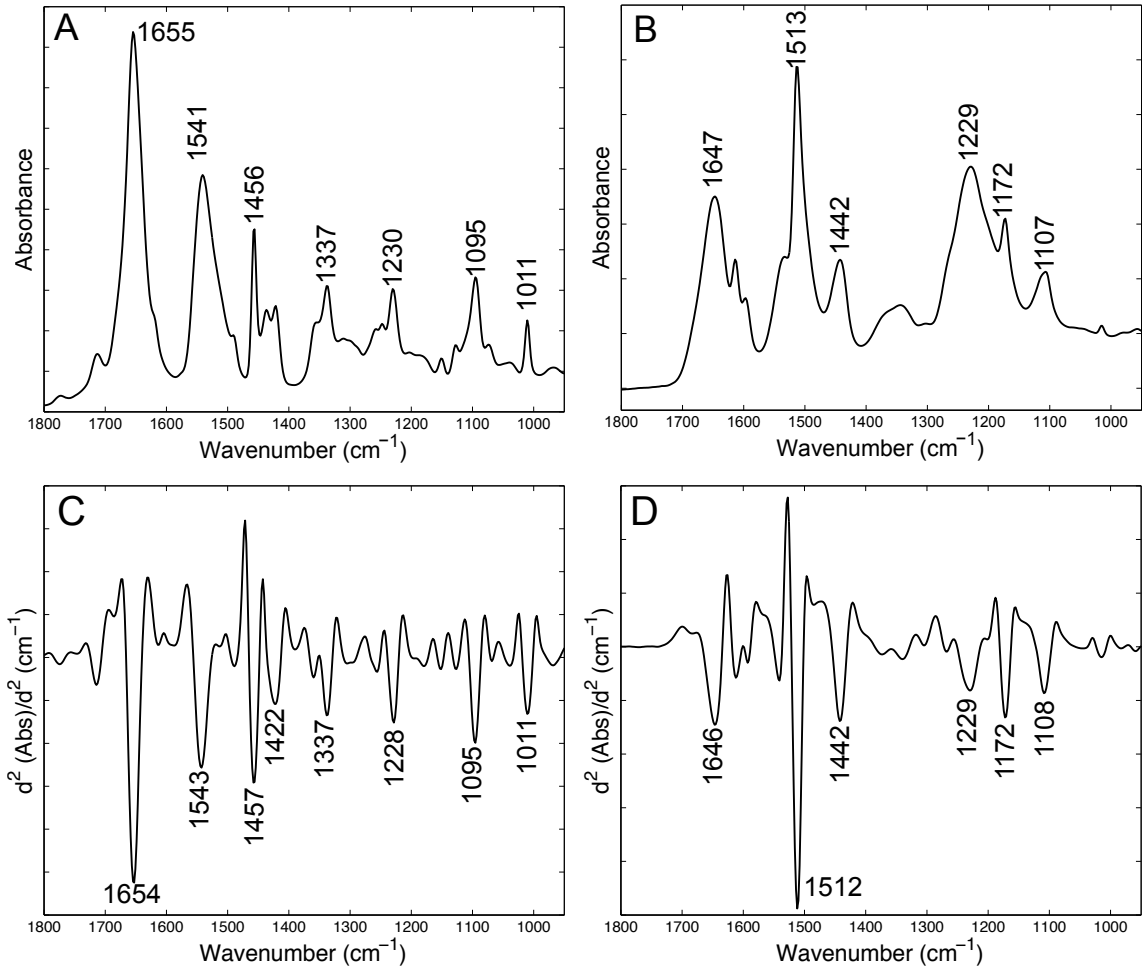


Figure 3.17: Model spectra of selected poly-amino acids. *(A) Solid poly-L-tryptophan. (B) Solid poly-L-tyrosine. (C) Second derivative of spectrum in A. (D) Second derivative of spectrum in B.*

3.4 Discussion and conclusions

The spectra presented here represent some, but not all, of the compounds that might be expected in human urine and tissue samples.

Human urine is a relatively simple mixture and most compounds in urine can be confidently identified and quantitated by comparison to model compound spectra (e.g. [49, 81, 82]). Since water makes up 95 % or more of urine it is usually necessary to remove the contribution of liquid water from sample spectra in order to extract information about the other components. After water, urea is the next most concentrated component in urine and can be identified by the absorbance band at 1468 cm^{-1} in solution spectra (Figure 3.1 A). The strong absorbance bands of urea between 1700 and 1550 cm^{-1} are also useful for confirming the urea assignment but residual water can influence the shape and intensity of these bands. Similarly, creatinine has a similar ab-

sorbance band pattern between 1700 and 1550 cm^{-1} making it difficult to discriminate between the two compounds using this region alone.

Creatinine has a strong absorbance band at 1492 cm^{-1} in solution, which shifts up to 1499 cm^{-1} in dried spectra (Figure 3.2 A & B). Although this band overlaps the 1468 cm^{-1} band of urea, the peak falls in a region that is free from absorbances of other major urinary components and would appear as a distinctive shoulder on the urea band. This is a useful feature as quantitation of creatinine is used clinically for normalising the concentrations of other urine constituents. The 1348 cm^{-1} band of creatinine, though lower intensity than the 1492 cm^{-1} band, also falls in a relatively clear region of the spectrum and could be used to confirm the presence of creatinine.

Phosphate (Figure 3.4) and sulphate (Figure 3.5) ions absorb predominantly between 1200 and 1000 cm^{-1} but can be distinguished by the additional band at 1157 cm^{-1} of phosphate ions and because the phosphate ion bands tend to be sharper. Glucose also absorbs in this region (Figure 3.6); urine from a healthy individual would not typically contain glucose but its presence could be determined by the unique absorbance pattern, with sharp peaks at 1150, 1107, 1080 and 1034 cm^{-1} in solution, and broad absorbances between 1500 and 1200 cm^{-1}.

All of the kidney stone materials presented have multiple sharp peaks. When diagnosing a kidney stone disease from IR spectra, several of these peaks would need to be present for positive identification, whilst absence of specific peaks could be used to eliminate compounds from the search. Stone materials are generally low concentration compared to the major urine components, however, since stone materials are insoluble they can be easily separated and concentrated by centrifuging the urine samples.

Tissue spectra are far more complex than urine spectra, containing many different types of proteins, carbohydrates, lipids and other metabolites. Spectra of diseased tissues are often characterised by subtle changes that cannot be assigned to specific compounds.

Global changes in protein content can be identified by changes in the position and shape of the amide I and amide II bands of protein at ~1650 and ~1550 cm^{-1}, respectively. However, there are many different types of protein in tissue and all have very similar IR spectra, meaning assignment to a specific protein is not generally possible.

The 1200 - 1000 cm^{-1} region of tissue IR spectra is particularly complicated, with contributions from carbohydrates/sugars (including glycoproteins), DNA, RNA and lipids. The absorbance bands of many of these compounds overlap significantly so that accurate assignment cannot be made, although many research groups have ten-

tatively assigned observed changes in tissue spectra to specific components, including, DNA/RNA [18, 31, 35, 55, 66, 88], mucin [65], collagen [90], cholesterol [20] and glycogen [18, 31, 35, 55, 65, 87].

Lipids, including phosphatidylcholine (Figure 3.11) and cholesterol (Figure 3.12), have strong sharp absorbances between 3000 and 2750 cm^{-1} and at around 1730 cm^{-1}, which are regions relatively free from contributions from other tissue components. These bands in tissue IR spectra can be attributed to lipids, but assignment to a specific type of lipid is more difficult.

Comparisons of model compounds spectra with biological sample spectra can be further complicated by the fact that the precise shape and position of absorbance bands can be influenced by factors such as hydration and the presence of other interacting compounds. This has already been illustrated by the differences seen between the solution and dried spectra and will be further discussed in results Chapter 4. Small shifts of 1-2 wavenumbers in the bands of compounds in mixed samples relative to pure solution are common but assignments may still be made in spite of such differences. However, correlating multiple peaks is often necessary for confident assignments, and often it is only possible to tentatively suggest a general type of compound that might be contributing to an IR spectrum.

Chapter 4

Hydration effects on infrared spectra

Author contributions: this chapter has been submitted to Applied Spectroscopy for publication. All data and figures in the publication were measured and prepared by Katherine Oliver, with contributions from Amandine Maréchal and Peter R. Rich for the 'simulated urine' experiment. The paper was written by Katherine Oliver with contributions from Amandine Maréchal and Peter R. Rich. Additional Figures have been included.

4.1 Introduction

There is a rapidly increasing literature on applications of vibrational infrared spectroscopy to analyses of complex biological tissues and fluids in order to identify distinguishing signatures of cell types, diseased tissues and diagnostic markers of specific diseases [48–50, 73, 81, 91, 114–117]. The mid-infrared (IR) spectra of such materials are inevitably complex combinations of many overlapping IR-active components. Diagnostic features of interest are in general minor components of these spectra and their small IR intensities are usually overlapped by much stronger bands from more dominant components.

Biological samples, particularly biofluids and thin tissue biopsy sections, are often dried before analysis [91, 116]. Drying removes the strong, broad absorbance bands of water and, in the case of measurements in attenuated total reflection (ATR) mode, increases the IR intensities of all other components by maximising their concentrations across the IR-active volume into which the evanescent wave penetrates. When applying analytical protocols to IR data obtained from dried and hydrated materials, it is important to take into account the fact that the IR spectra of many materials are different in aqueous solution versus the fully dried states. However, a further complication is that some materials can also exhibit additional distinct states during drying whose IR characteristics differ from both the solution and fully dried forms. This can be an important additional consideration when analysing materials in which residual water levels are not easy to remove in a consistent manner since such intermediate states can become trapped within the material as it dries.

Urea, with its strong interactions with water, is a particularly clear example of such behaviour. Its strong hydrogen bonding propensity results in its well-known chaotropic effects in aqueous solutions [118]. Molecular modelling of its interaction with an increasing number of water molecules suggests that the first four water molecules form the primary hydration sphere and three more form the secondary hydration sphere [119–121]. Additional types of urea-water complexes are also predicted [119]. Such

urea-bound water molecules have been detected directly in aqueous solutions of urea by polarization-resolved mid-IR pump-probe spectroscopy [122]; bulk water molecules beyond the hydration sphere around urea are not significantly perturbed [120, 122–124], though this has been questioned with NMR studies of long range order [123]. At high concentration, urea may also form dimers or higher order structures [119, 121, 125, 126], although several studies have questioned their existence [124, 127, 128]. Some of these controversies may arise due to the different nature of the probe between different spectroscopic techniques. For an extensive review see ref. [128].

Here, IR spectroscopy is used to demonstrate the different forms of urea that can form at different hydration states and how these influence the infrared spectra of wet and dried urine samples. IR analyses of creatinine and dehydration profiles of other urinary constituents are also described to further illustrate the need to consider hydration state when interpreting IR data of biological samples.

Urea is an important breakdown product of nitrogen-containing compounds in the body, such as dietary and muscle protein. It provides a safe way for the body to excrete nitrogen as it is highly soluble in water and non-toxic. Creatinine is a byproduct of muscle metabolism, with its production in the body being proportional to body mass [106]. Urea and creatinine are two major components of urine that are often measured clinically (by non-IR methods) and can be used as markers for urine concentration against which other components can be normalised. Creatinine can also be used as an indicator of renal function and is used to calculate the glomerular filtration rate. Several methods for quantitation of urea and creatinine in wet urine samples have been described that are based on their well-defined solution IR spectra [49, 81]. For example, as described here, urea concentrations in aqueous solution or in wet urine can be quantitated by fitting a Gaussian component to its strong v_{as} (asymmetric CN stretch) band at 1468 cm^{-1} [109]. When urea is completely dry this v_{as}(CN) band shifts to 1464 cm^{-1}. However, in dried urine samples, the position of the v_{as}(CN) band of urea does not correspond to either the wet or dried forms of pure urea (Figure 4.1).

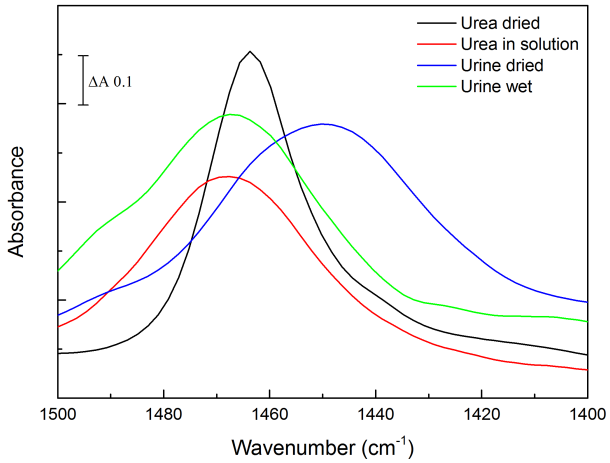


Figure 4.1: Variation in the v_{as}(CN) band of urea in wet and dried pure urea and urine spectra. *The shoulder at around 1492 cm^{-1} in the wet and dried urine spectra is due to creatinine.*

It was found that at least two additional states of urea with a v_{as}(CN) band at 1454 or 1443 cm^{-1} can be observed in partially-hydrated urea. Similar, though less dramatic, hydration effects were observed with creatinine and at least one intermediate form was identified in partially-hydrated samples. When whole urine is dried, the solid materials tend to trap these intermediate forms. Hence, accurate protocols for component deconvolution, or for comparison of wet and dried biological fluids and tissues, should accommodate such additional states.

4.2 Transmission IR spectra of urea

Spectra recorded in ATR mode can be influenced by the refractive index of the sample. To ensure that the observed change in urea v_{as}(CN) band position and shape was not simply due to a difference in refractive index of wet *versus* dried samples, spectra of urea were recorded in transmission mode. The transmission mode spectra of a pure urea solution dried and partially rehydrated show the same shifts as those seen in the ATR spectra, with the v_{as}(CN) band centred at 1465 cm^{-1} in the dried spectrum (1 cm^{-1} higher than observed in ATR spectra) and shifting to 1454 cm^{-1} in the partially hydrated spectrum (Figure 4.2). This suggests that the changes in the spectra between wet and dried pure urea and urine samples are not the result of a difference of refractive index.

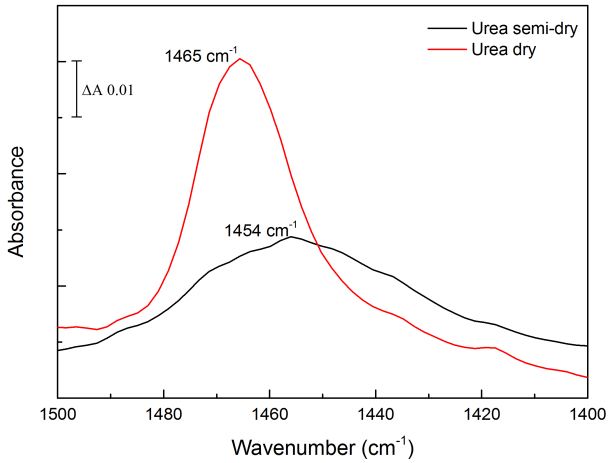


Figure 4.2: Transmission spectra of urea. *8 µL of a 50 mM solution of urea was dried on to a 2 mm thick CaF_2 window. A dried spectrum was recorded (red trace) before the sample was rehydrated by placing drops of water around the dried urea spot and a second spectrum recorded (black trace).*

4.3 Kinetic resolution of multiple urea hydration states

As expected with a small molecule that forms multiple hydrogen bonds with water, the majority of bands of urea change their frequencies and relative intensities between the solution and fully dried states. Dehydration resulted in loss of the underlying broad absorbance bands of liquid water in the 1700-1600 cm^{-1} and 3700-2900 cm^{-1} ranges, with concurrent intensification of narrower bands of urea in these regions (Figure 4.3 A & B). The rehydrated spectrum is very similar to the solution urea spectrum and the fully dried spectrum is essentially the same as that of pure, crystalline urea (Figure 4.4 A & B) [109, 110]. However, the behaviour was clearly not a simple transition between these two states. For example, the υ_{as}(CN) band in the rehydrated spectrum was centred at 1466 cm^{-1}, close to its position at 1468 cm^{-1} in the spectra of urea in solution at physiological concentrations. As the sample became dehydrated this peak gradually downshifted, transiently reaching a minimum at 1448 cm^{-1} before rapidly upshifting to the 1464 cm^{-1} value characteristic of the fully dried state (Figure 4.3 C & D).

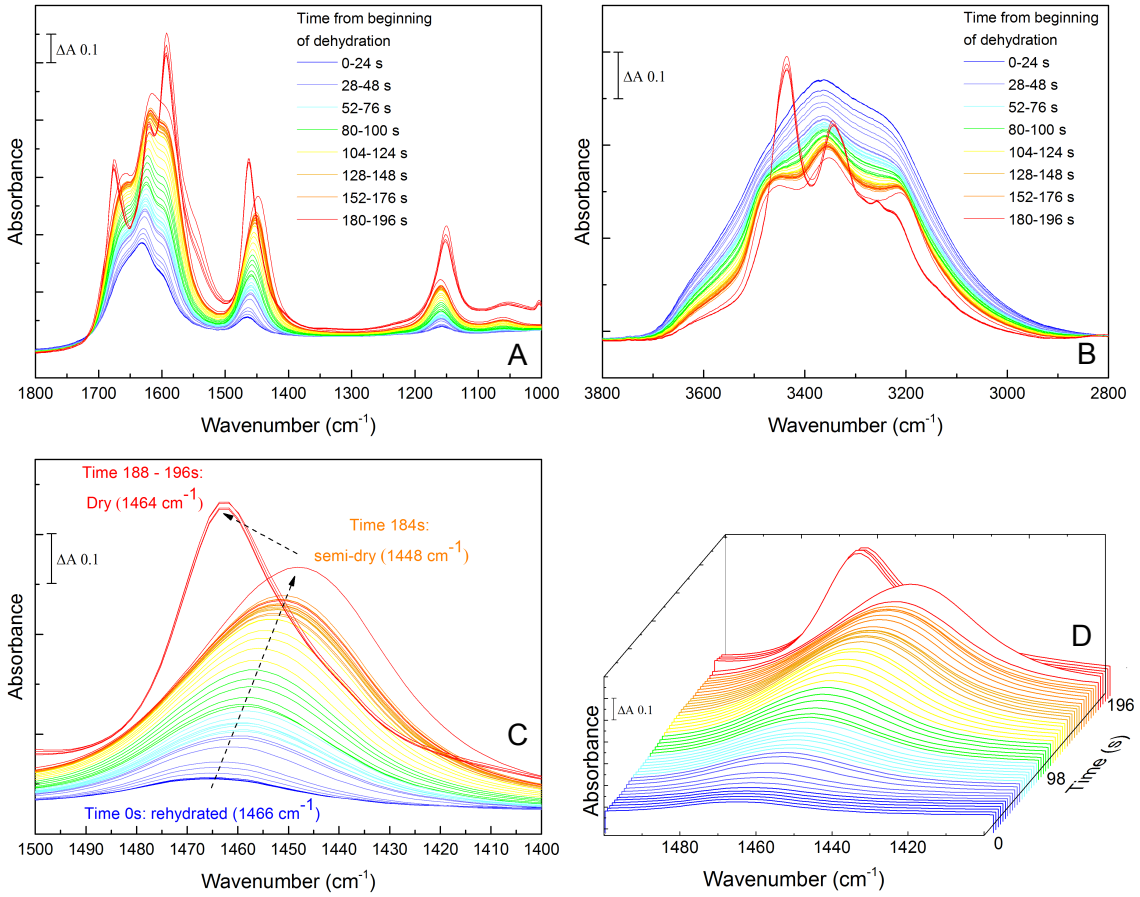


Figure 4.3: Serial spectra of a rehydrated urea sample during drying. *3 μL of a 50 mM solution of urea were firstly dried on the prism and rehydrated in a water-saturated atmosphere. Spectra were then recorded during drying over an approx. three minute period. (A) the 1800- 1000 cm^{-1} 'fingerprint' region; (B) the 3800-2800 cm^{-1} region. (C and D) 2D and 3D plots of the 1500–1400 cm^{-1} region showing the υ_{as}(CN) peak behaviour. Rehydrated spectra are shown in blue and dried spectra in red.*

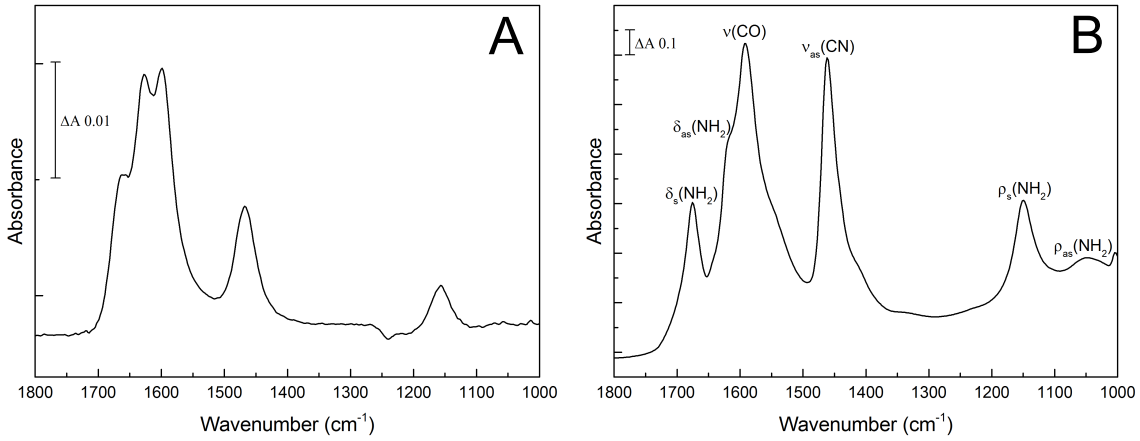


Figure 4.4: Solution and crystalline spectra of urea. *(A) spectrum of 200mM pure urea in solution after subtraction of pure water and water vapour spectra. (B) crystalline urea with band assignments.*

4.4 Stepwise rehydration of urea

By controlling the relative humidity of the environment surrounding the sample, it was possible to stabilise the intermediate partially-hydrated forms of urea. For example, Figure 4.5 summarises the peak positions of the υ_{as}(CN) band after stabilisation at different relative humidities. As the humidity decreased, the peak at 1468 cm^{-1} in solution spectra downshifted to an intermediate state of 1456 cm^{-1} at 79 % humidity. However, as the humidity was further decreased, the peak then upshifted towards the fully dried value of 1464 cm^{-1}.

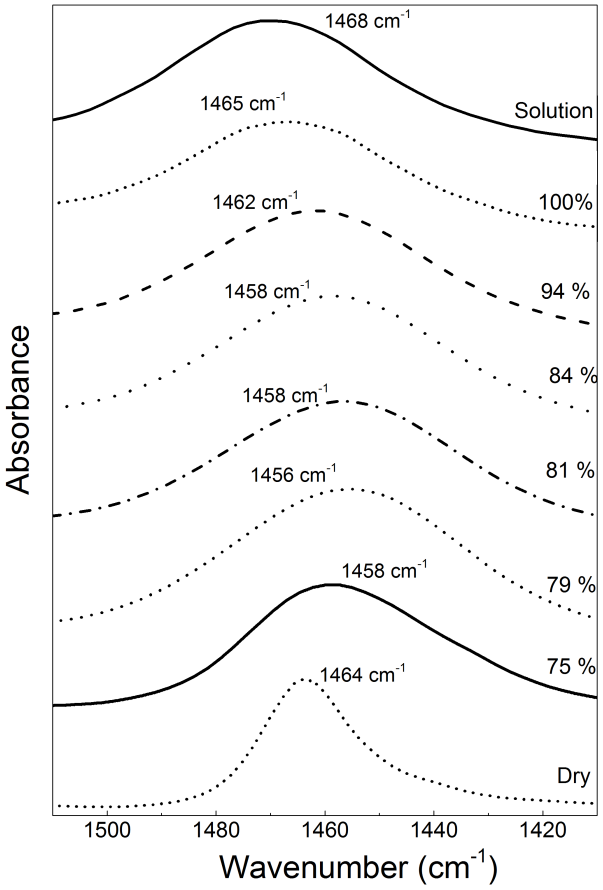


Figure 4.5: Spectra of the v_{as}(CN) band of urea between 1510 and 1410 cm^{-1} at fixed relative humidities. *The v_{as}C-N band peak at 1468 cm^{-1} in solution downshifts as relative humidity decreases to 79 %, before upshifting towards the 1464 cm^{-1} position characteristic of dry urea at lower relative humidities (spectra are not shown to scale).*

4.5 Component fitting to urea spectra

Combinations of Gaussian and pseudo-Voigt components were fitted to the v_{as}(CN) urea peak of typical spectra obtained between 1510 and 1410 cm^{-1} as the relative contributions of hydrated, partially-hydrated and fully dried forms varied during drying. Before fitting, a correction was made for the ramping of spectral intensities caused by recording in ATR mode (see Methods). Four components (Table 4.1) were required to adequately fit all spectra in this frequency range based on the root mean square error compared to a three of fewer component fit. Fitting with additional components did not improve the error of the fit.

Table 4.1: Parameters of components required for simulation of urea spectra between 1510 and 1410 cm^{-1} corresponding to the v_{as}(CN) band. *[92]*

Component centre of the urea v_{as}(CN) band (cm^{-1})	**Function type**	**m_u(fraction of Lorentzian component)**	**FWHM (cm^{-1})**
1468	Gaussian	n/a	32
1464	Pseudo-Voigt	0.61	19
1454	Pseudo-Voigt	0.66	36
1443	Pseudo-Voigt	0.97	41

The fitting data with these components are shown in Figure 4.6. For reference, the best fit of combinations of these same components to this region of urea in solution is shown in Figure 4.7; this is dominated by the Gaussian component centred at 1468 cm^{-1} that also dominates the rehydrated spectrum (Figure 4.6 A). As the rehydrated sample dried, the 1468 cm^{-1} component diminished and components at 1454 and 1443 cm^{-1} emerged (Figure 4.6 B). As the drying progressed, the 1443 cm^{-1} component became the dominant species (Figure 4.6 C) before the band rapidly upshifted to the 1464 cm^{-1} form that is characteristic of fully dried urea (Figure 4.6 D).

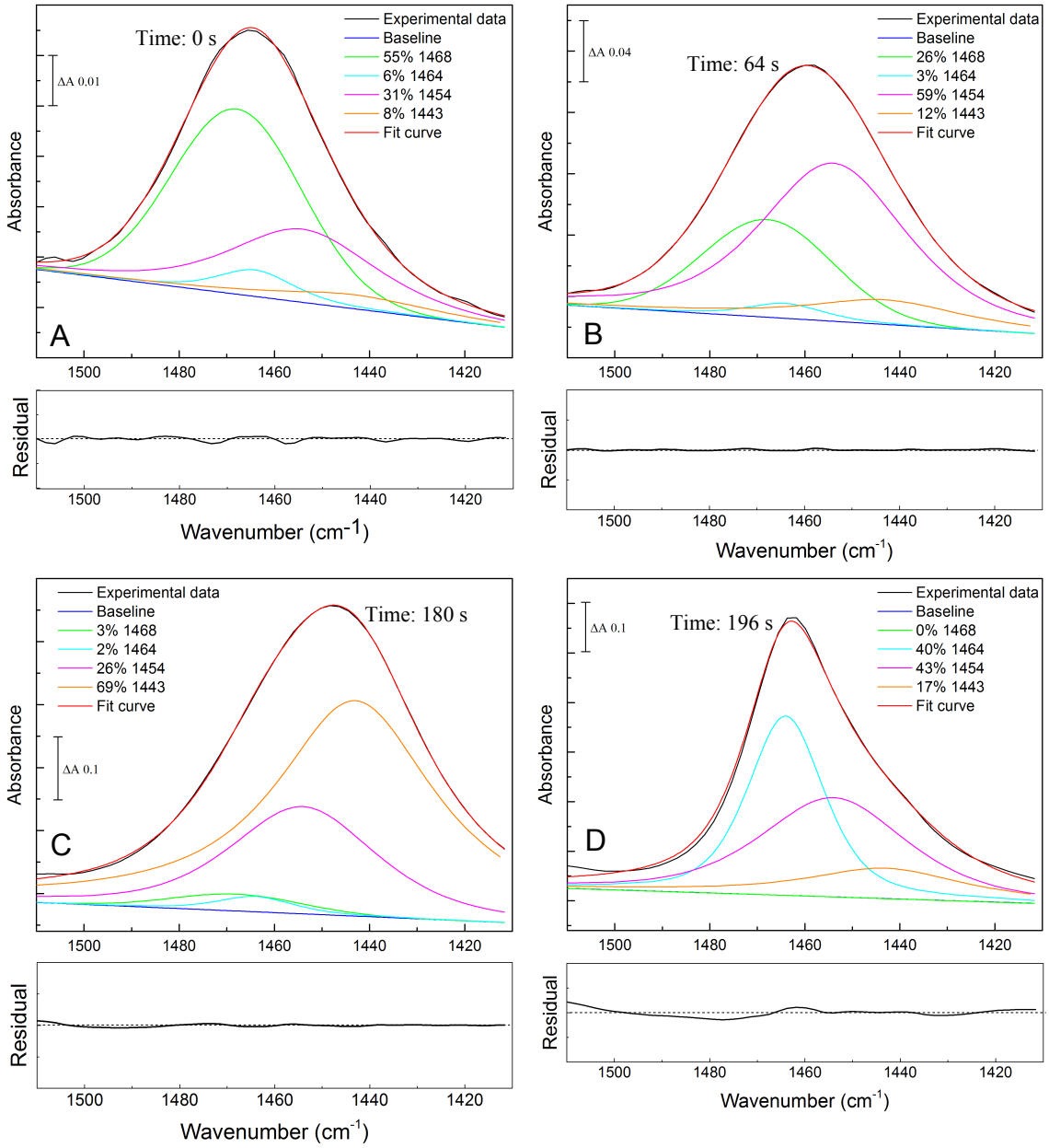


Figure 4.6: Component fitting to the υ_{as}(CN) band of urea during the transition from rehydrated to dry states. *Four spectra were selected from the series obtained during drying of rehydrated urea of Figure 4.3. Times from beginning of the dehydration are indicated on each panel. Before fitting, spectra were corrected for effects of ramping caused by recording in ATR recording mode, followed by optimal fitting of the intensities of the four components of fixed frequencies and bandwidths (Table 4.1), together with unrestricted linear sloping baselines. (A) rehydrated; (B and C) intermediate partially-hydrated; (D) dry states. The residuals for each overall fit are shown underneath each panel at the same scale as the corresponding component fit, with the centre dotted line representing zero.*

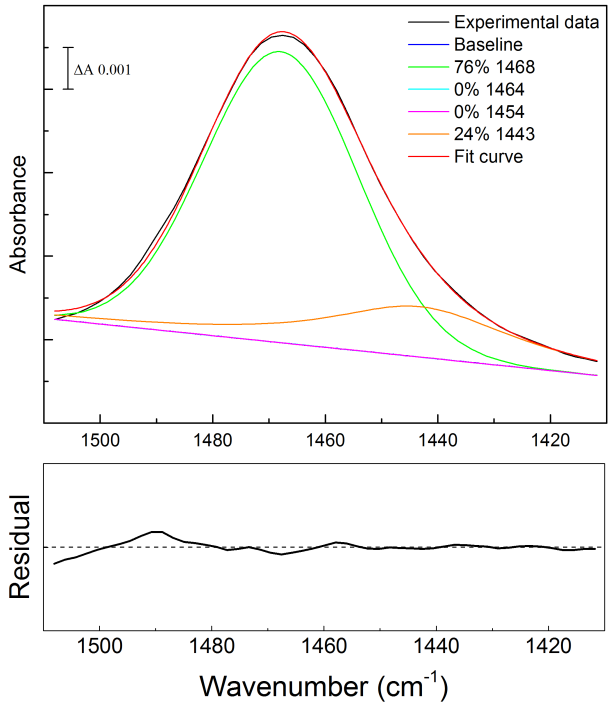


Figure 4.7: Component fitting to the v_{as}(CN) band of urea in solution. *A spectrum of 200 mM urea in aqueous solution was recorded. Absorbance contributions of bulk water were firstly subtracted using an ATR spectrum of pure water and the 1510–1410 cm^{-1} region was then corrected for the effects of ramping caused by recording in ATR recording mode. An optimal fit was then made of intensities of the four components of fixed frequencies and bandwidths (Table 4.1), together with an unrestricted linear sloping baseline. Fractions of each component were calculated from the peak areas (integrated intensities). The residual for the overall fit is shown underneath at the same scale as the component fit, with the centre dotted line representing zero.*

4.6 Hydration states of creatinine

Hydration-dependent variations in absorbance band frequencies and intensities occur with other urinary components, though none that were studied were as marked as those seen with urea. Creatinine is a second major urinary component and has an absorbance band at 1492 cm^{-1} in solution that partly overlaps with the v_{as}(CN) absorbance band of urea. Based on previously published work [129, 130] and frequency predictions performed in Gaussian 09 [101] of pure creatinine in a vacuum (dielectric constant = 1.0) matrix, this creatinine band can be assigned to a single complex normal mode (v(C=N) (30) v(CN) (27) δ(NCH) (24)). Spectra were recorded during drying and components were again fitted to the data between 1520 and 1470 cm^{-1} after correction for the ramping effect of recording in ATR mode. The 1492 cm^{-1} band upshifted during dehydration to an intermediate position at 1501 cm^{-1} (Figure 4.8). Continued drying resulted in a slight

downshift to 1499 cm^{-1}, the same position as the band in dry, crystalline creatinine (not shown). In both solution and rehydrated states, this band can be approximately described by a single Lorentzian component centred at 1492 cm^{-1} with FWHM of 24 cm^{-1} (Figure 4.9 A and Table 4.2). After drying, or in crystalline creatinine, the band approximates to a single Lorentzian component at 1499 cm^{-1} with FWHM of 29 cm^{-1} (Figure 4.9 D). However, an additional Gaussian component at 1501 cm^{-1} with FWHM of 23 cm^{-1} is required in order to adequately simulate all spectra obtained in the partially hydrated conditions arising between these extremes (Figure 4.9 B & C). All spectra could be adequately represented by a combination of these three components.

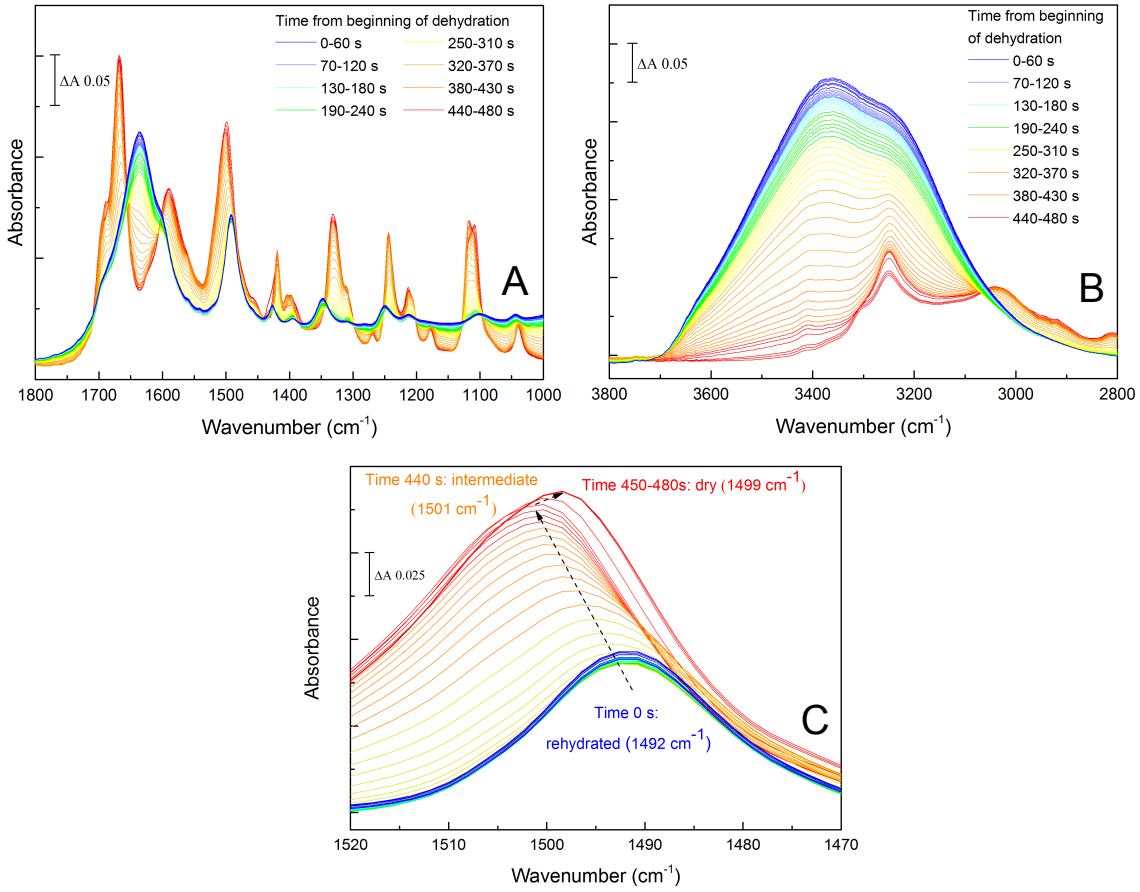


Figure 4.8: Serial spectra of a rehydrated creatinine sample during drying. *3 µL of a 10 mM solution of creatinine was firstly dried onto the prism surface and rehydrated in a water-saturated atmosphere. Spectra were then recorded during drying over a three minute period. (A) the 1800- 1000 cm^{-1} 'fingerprint' region; (B) the 3800-2800 cm^{-1} region. (C) the 1500–1400 cm^{-1} region showing the (υ(C=N) (30) υ(C-N) (27) δ(NCH) (24)) band behaviour. Rehydrated spectra are shown in blue and dried spectra in red.*

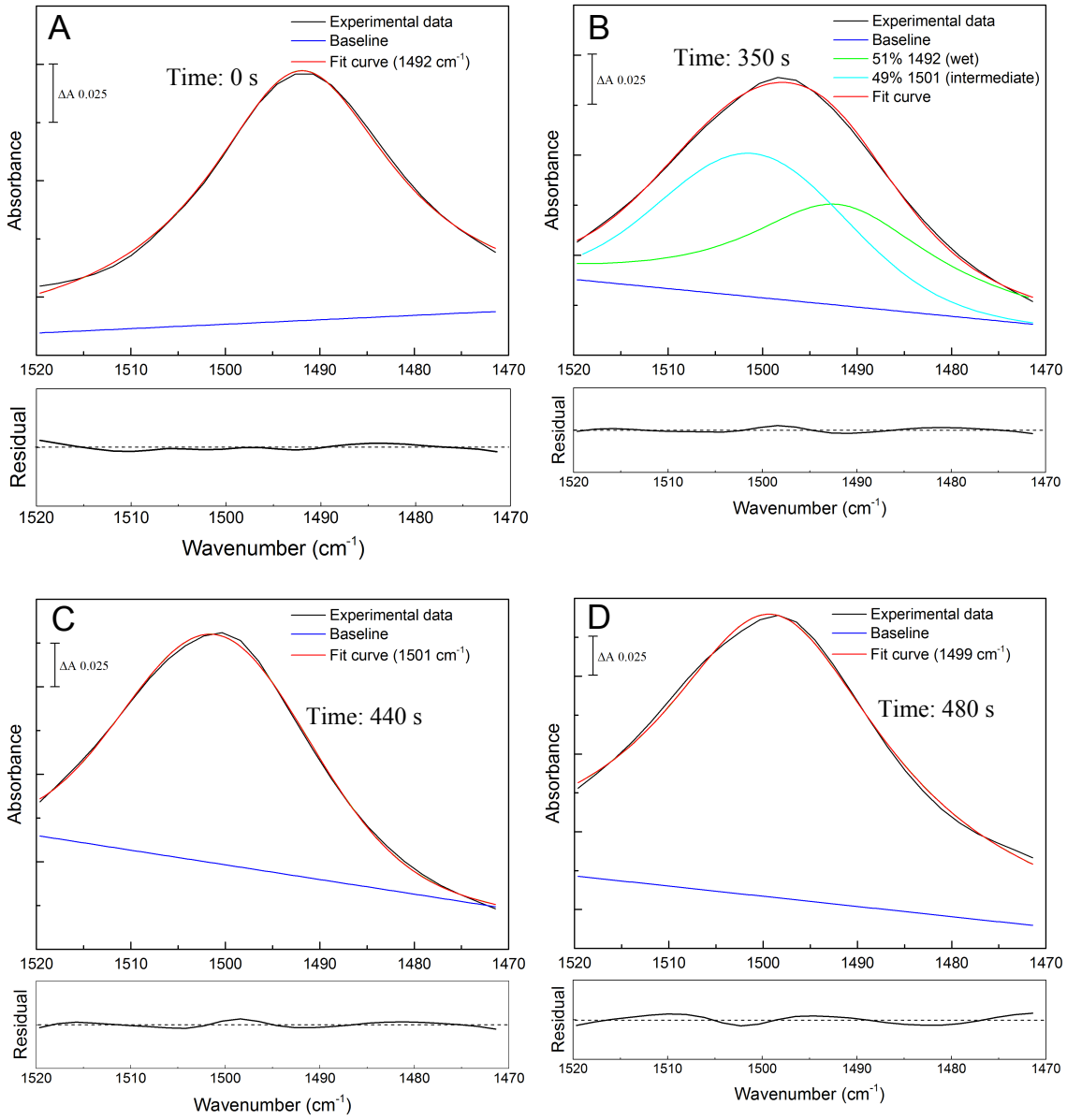


Figure 4.9: Component fitting to the 1520 and 1470 cm^{-1} spectral region of creatinine during the transition from rehydrated to dry states. *3 μL of a 10 mM solution of creatinine were firstly dried onto the prism surface and then rehydrated in a water-saturated atmosphere. Four spectra were selected from the series obtained during drying of this rehydrated sample. Spectra shown were corrected for effects of ramping caused by recording in ATR recording mode, followed by optimal fitting of the intensities of the three components of fixed frequencies and bandwidths (Table 4.2), together with unrestricted linear sloping baselines. (A) rehydrated; (B and C) intermediate partially-hydrated; (D) dry states. The residuals for each overall fit are shown underneath each panel at the same scale as the corresponding component fit, with the centre dotted line representing zero.*

Table 4.2: Parameters of components required for simulation of creatinine spectra between 1520 and 1470 cm^{-1} corresponding to the (υ(C=N) (30) υ(CN) (27) δ(NCH) (24)) band.

Component centre of (υ(C=N) (30) υ(C-N) (27) δ(NCH) (24)) band (cm^{-1})	**Function type**	**FWHM (cm^{-1})**
1492	Lorentzian	24
1499	Lorentzian	29
1501	Gaussian	23

4.7 Hydrated forms of urea in dried urine samples

The dehydration protocol was repeated with urine samples from healthy donors. Figure 4.10 shows how the spectra of two typical urine samples varies during drying.

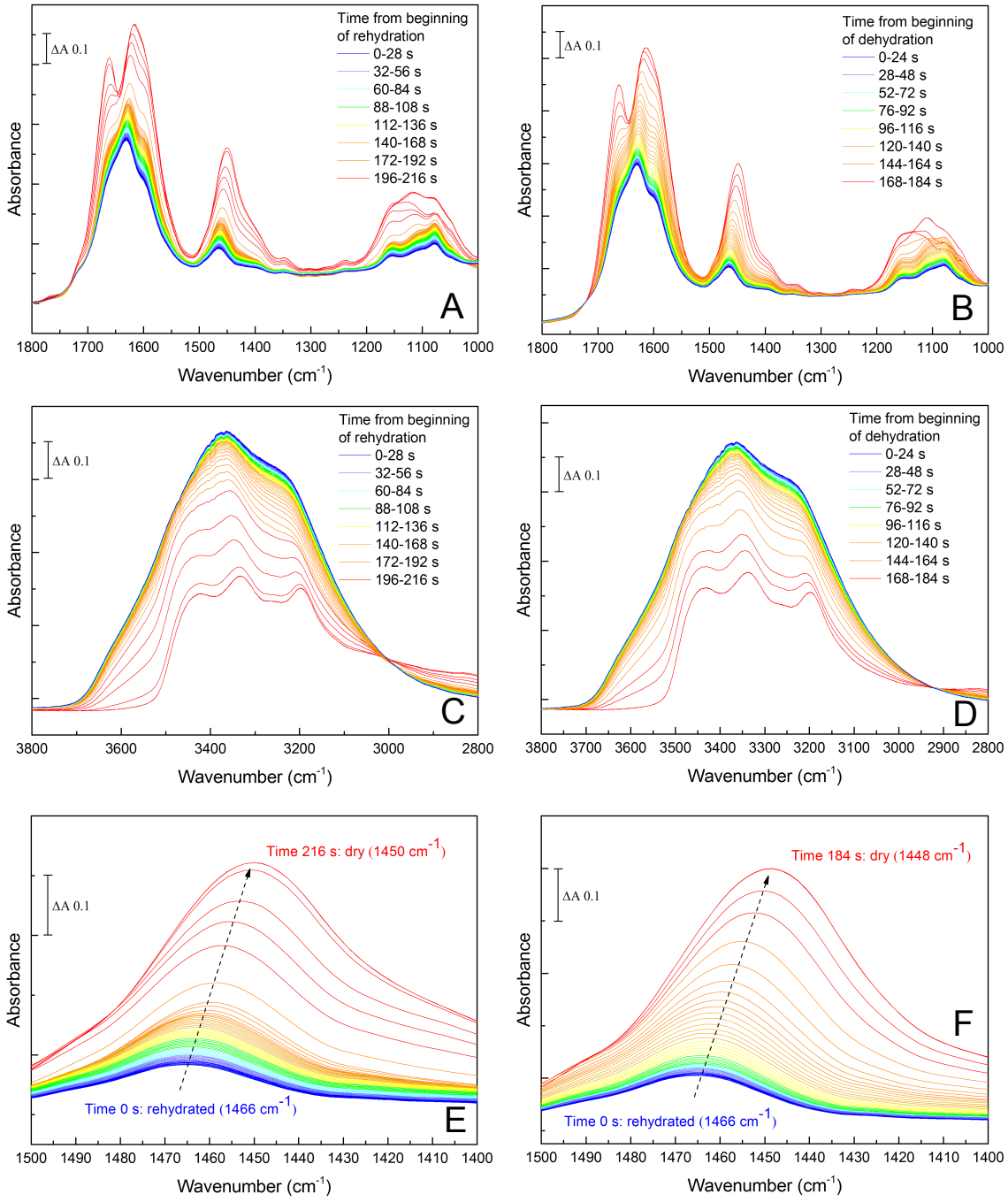


Figure 4.10: Serial spectra of rehydrated urine samples during drying. *3 μL of 1 in 4 diluted urine samples were firstly dried onto the prism surface and rehydrated in a water-saturated atmosphere. Spectra were then recorded during drying over a three minute period. (A and B) the 1800-1000 cm^{-1} 'fingerprint' regions samples N1 (A) and N2 (B); (C and D) the 3800-2800 cm^{-1} regions. (E and F) the 1500–1400 cm^{-1} regions. Rehydrated spectra are shown in blue and dried spectra in red.*

Figure 4.11 shows the dehydration spectra of the υ_{as}(CN) band of urea for four additional urine samples.

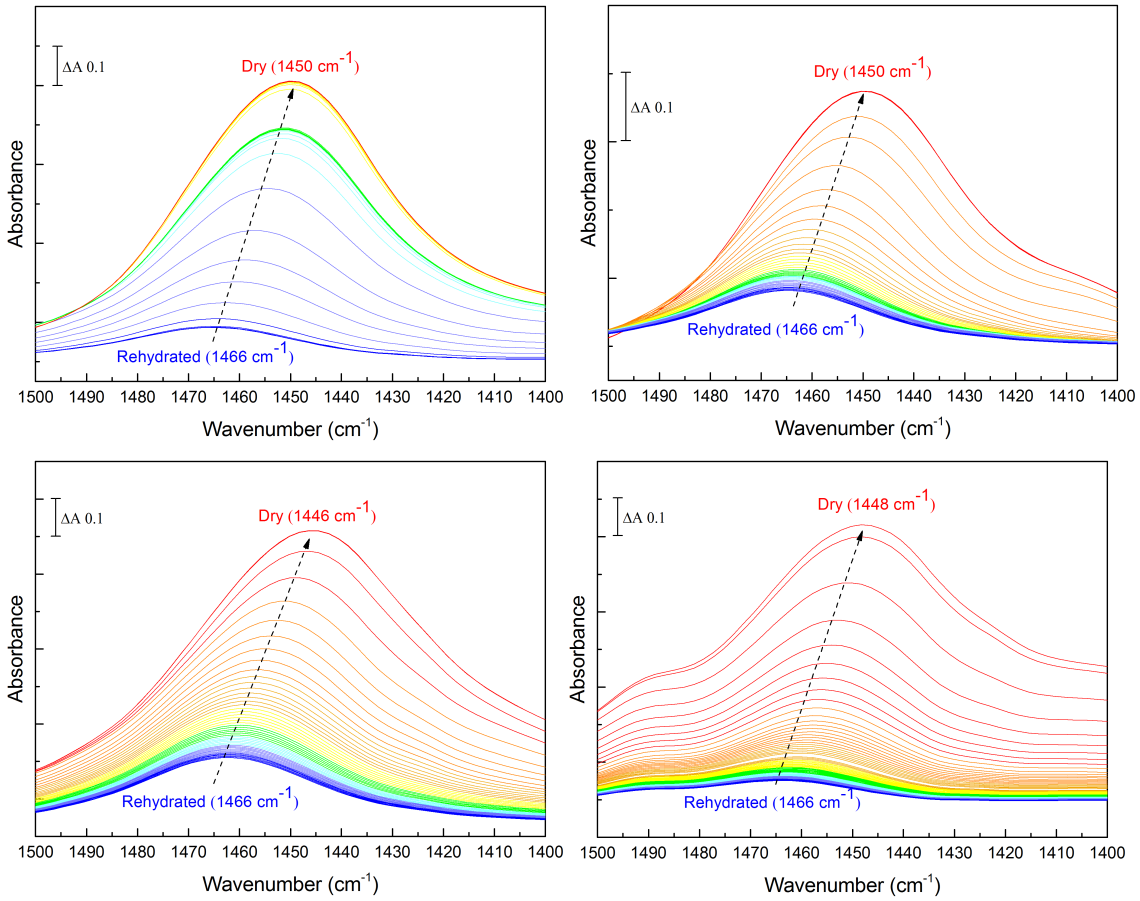


Figure 4.11: Serial spectra of the υ_{as}(CN) band of urea in four additional urine samples during drying. *3 μL of 1 in 4 diluted urine samples were firstly dried onto the prism surface and rehydrated in a water-saturated atmosphere. Spectra were then recorded during drying over a three minute period.*

The 1510-1410 cm^{-1} region of spectra rehydrated human urine samples is dominated by the υ_{as}(CN) band of urea, together with a smaller contribution from the 1492 cm^{-1} normal mode of creatinine. Spectra were analysed by fitting the four components of pure urea (Table 4.1) together with additional smaller contributions from the possible forms of creatinine (Table 4.2) and unrestricted linear sloping baselines (Figure 4.12).

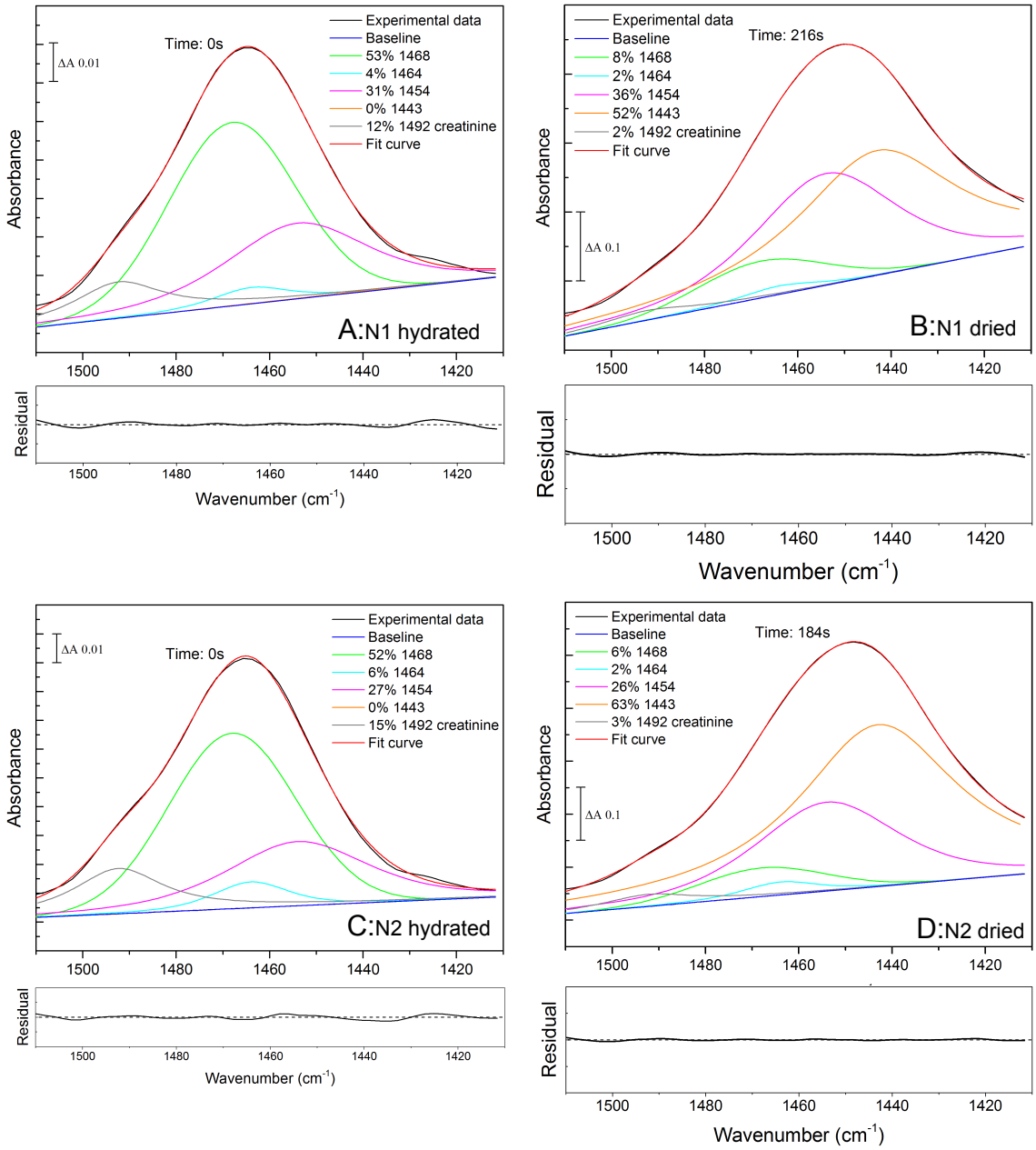


Figure 4.12: Fitting of urea and creatinine components to the 1510–1410 cm^{-1} region of spectra of human urine. *3 μL samples of two typical healthy human urine samples were dried onto the prism surface and rehydrated with a water-saturated atmosphere. Spectra were recorded after stabilisation ((A) rehydrated sample N1; (B) rehydrated sample N2). Samples were then dried and spectra were rerecorded after they had stabilised ((C) dried sample N1; (D) dried sample N2). The fractional contributions of each component were determined from integrals of peak areas. The residuals for each overall fit are shown underneath each panel at the same scale as the corresponding component fit, with the centre dotted line representing zero.*

In the rehydrated state, as expected, the spectrum is dominated by the 1468 cm^{-1} form of urea together with a 10-15% contribution from the 1492 cm^{-1} form of creatinine (Figure 4.12 A & C). This is similar to that observed with fresh liquid urine

(Figure 4.13).

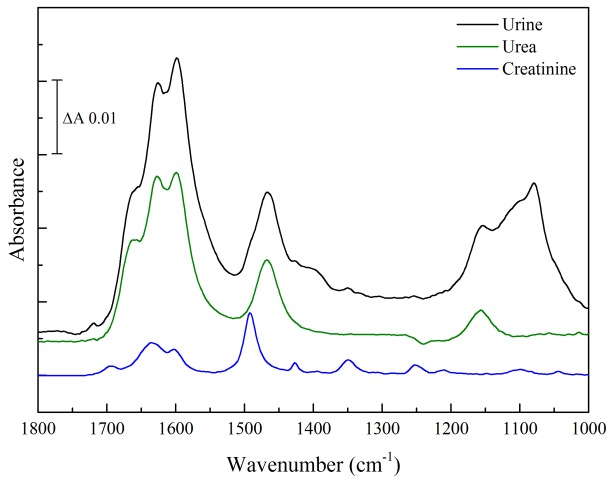


Figure 4.13: ATR-FTIR spectrum of a typical urine sample. *Also shown are the solution spectra of urea and creatinine. Samples were recorded wet and pure water and water vapour spectra were subtracted from the sample spectra.*

However, even after drying by extensive exposure to the dry nitrogen flow, the urea components (Figure 4.12 B & D) corresponded to those of its partially-hydrated (1454 and 1443 cm^{-1}) states (Figure 4.6 C), rather than the 1464 cm^{-1} band of the fully dried form. Because the enhancement of the urea band intensity on drying is much greater than that of creatinine (compare Figures 4.3 and 4.8), the creatinine contributions to this region of the dried spectra become much smaller (< 3%). In Figure 4.12 B and 4.12 D, a single component at 1492 cm^{-1}, corresponding to hydrated creatinine, appears adequate, consistent with water retention in the dried urine samples, though accurate assignment to its different states cannot be made with confidence because of its very low relative intensity.

The retention of intermediate hydration states of urea in dried urine samples is also clearly seen in the 3700 - 2900 cm^{-1} region of spectra of dried urine (Figure 4.10 C & D). Whereas it is clear that the broad band of condensed liquid water has been lost, indicating that no bulk liquid water remained, the remaining urea bands resembled those of the partially-hydrated forms, rather than those of the fully dried states (Figure 4.6 D). Hence, again, it is clear that drying had trapped predominantly the intermediate states of urea, rather than the fully dried form.

4.8 Factors in urine affecting observed behaviour

The question arises as to what factor(s) control the different behaviours of urea when dried in pure solution versus urine. A 'simulated urine' solution was prepared containing 115 mM urea, 50 mM NaCl, 29 mM NH_4Cl, 19 mM KH_2PO_4, 10 mM K_2SO_4 and 5 mM creatinine, buffered to pH 6 with potassium hydroxide [106]. When a 3 μL sample was dried onto the ATR prism, the υ_{as}(CN) band of urea stabilised around 1445 cm^{-1} (Figure 4.14), reproducing the behaviour of urea in urine. However, the relative amount of the 1464 cm^{-1} component formed on drying increased when one of the major components (ammonium chloride, creatinine or phosphate) was omitted, or if the mixture was diluted substantially before drying a 3 μL sample. In contrast, if one component was omitted but a replacement solute was added to maintain the solute mass, then the urea again failed to form the 1464 cm^{-1} component. Furthermore, if urine samples were extensively diluted, then a significant fraction of the 1464 cm^{-1} form of urea appeared on drying. It is therefore concluded that the urinary effect is a non-specific one, caused simply by the mass of dried components (in urine, urea accounts for only 25-50% of total solutes) creating a surface barrier that prevents complete dehydration of the material close to the prism surface.

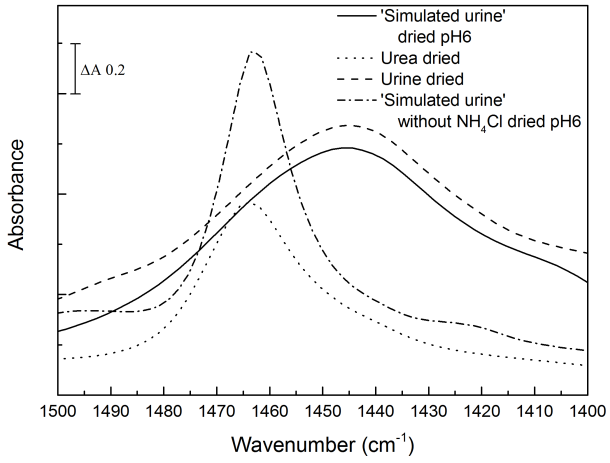


Figure 4.14: Behaviour of the υ_{as}(CN) urea band in dried 'simulated urine'. *3 μL of 'simulated urine' containing 115 mM urea, 50 mM NaCl, 29 mM NH_4Cl, 19 mM KH_2PO_4 , 10 mM K_2SO_4 and 5 mM creatinine were dried onto the prism surface and the spectrum was recorded after stabilisation (solid line). Also shown are spectra of 3 μL 50 mM urea after drying (dotted line) and 3 μL healthy donor urine after drying (dashed).*

4.9 Electronic structure modelling and physical basis of intermediates

Changes in the spectra of pure urea during dehydration must arise from changes in urea-water and/or urea-urea interactions. A similar phenomenon has already been reported by Grdadolnik and Maréchal [120] and Jung *et al. [125]* who observed a shift in the urea v_{as}(CN) band in very concentrated solutions; these were assigned to specific urea-urea and/or urea-water interactions in the limited water environments. In aqueous solution, urea interacts directly with 4 water molecules [119–121]. Gaussian 09 software was used to calculate the IR spectra of urea alone, in different hydration states and in interactions with other urea molecules. The predicted frequency of the v_{as}(CN) band (Table 4.3) downshifted from 1470 to 1381 cm^{-1} as the number of surrounding water molecules decreased from 4 to 0. This downshift was due to the loss of H-bonding interactions, in particular between water hydroxyls and the carbonyl oxygen of urea [120]. In support of this, the most dramatic predicted changes corresponded to the removal of the last two waters which, in these simulations, were those interacting with the carbonyl group. Conversely, the predicted frequency of the v_{as}(CN) band of anhydrous urea upshifted from 1386 to 1445 cm^{-1} as its carbonyl oxygen made an increasing number of H-bonds with adjacent urea $-NH_2$ group(s) (Table 4.3). Although multiple v_{as}(CN) frequencies were predicted (one for each urea molecule), the highest wavenumber correlated with those ureas with the greatest number of H-bonding interactions of their carbonyl oxygen. For example, the 1381 cm^{-1} band in isolated urea upshifted to 1470 cm^{-1} in a urea with three H bonds in the 5 urea ensemble. Although the precise peak positions predicted during the simulation vary from those observed in the experimental data, these calculations, at least qualitatively, support a proposal that the experimentally observed initial frequency downshift of the v_{as}(CN) band on drying is due to dehydration of urea, and the subsequent upshift arises from direct urea-urea interactions as the last waters are removed.

Table 4.3: Predicted frequencies of υ_{as}(CN) band of urea in combination with water or other urea molecules. *Calculations were performed with Gaussian 09. Where larger urea ensembles were modelled, only the frequencies of those ureas with the highest number of H-bonds to their carbonyl oxygen are listed (see text for details).*

Structure	Number of H-bonds to the carbonyl oxygen	Wavenumber (cm^{-1})
1 urea + 4 H_2O	2	1470
1 urea + 2 H_2O	2	1462
1 urea + 1 H_2O	1	1420
1 urea	0	1381
2 ureas	1	1410, 1428
3 ureas	2	1419, 1439
5 ureas	3	1470

4.10 Hydration effects of other compounds in urine

When analysing spectra of biological material for changes associated with disease, it is often the subtle spectral changes that are important. As has been demonstrated with urea and creatinine, it is important to consider environmental humidity when analysing spectra of dried or partially-dried samples. It is useful to be aware of the IR dehydration profiles of the sample constituents to avoid mistaking a hydration state change as a significant biological marker of disease. Since urine samples are often analysed by IR spectroscopy in the dried state the dehydration profiles of the main urine constituents have been recorded, along with a few disease biomarkers. For each compound the 'fingerprint' region has been plotted along with the 3800 - 2800 cm^{-1} region to show how the broad water band changes during dehydration.

4.10.1 Ammonium chloride

Ammonium ions have a buffering role in human urine and can be excreted in response to a drop in pH. Ammonium ions in urine are predominantly generated in the cells of the proximal tubule of the kidney, primarily from the metabolism of the amino acid, glutamine [111]. Ammonium ions, like urea, interacts strongly with water so some spectral changes are seen during dehydration (Figure 4.15), in particular a strong

absorbance band appears at around 1410 cm^{-1} in the dried spectrum that is present only as a subtle shoulder in the hydrated spectrum. The intensity of the band at around 1290 cm^{-1} also increases significantly during drying. There are small downshifts in the position of these absorbances bands but the changes are not as remarkable as those seen for urea. Since the sample represented in Figure 4.15 was buffered with phosphate buffer, some of the changes occurring in the 1200 - 1000 cm^{-1} region can be attributed to phosphate. Comparison of Figure 4.15 with Figure 4.17, which shows the dehydration spectra of phosphate, can be used to identify which spectral changes can be attributed to ammonium chloride.

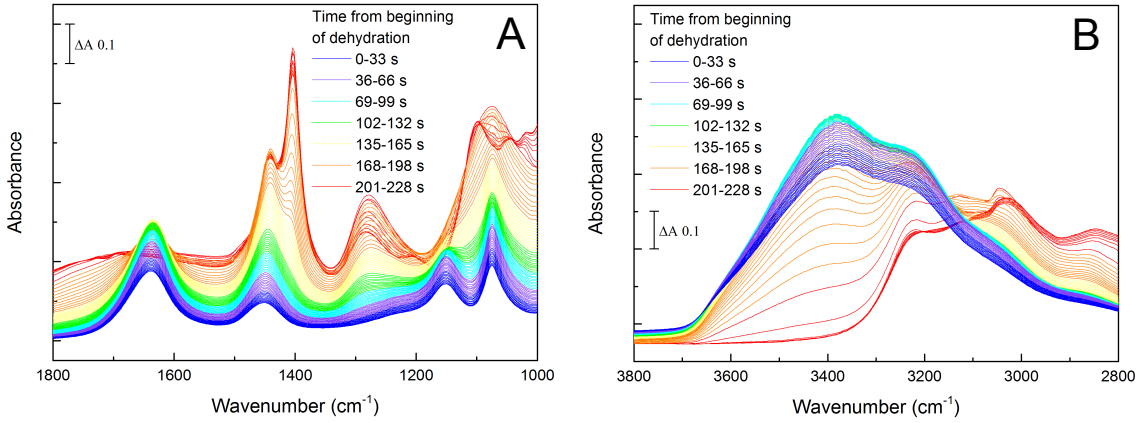


Figure 4.15: Serial spectra of a rehydrated ammonium chloride sample during drying. *3 μL of a 29 mM solution of ammonium chloride (adjusted to pH 6 with phosphate buffer) was firstly dried on the prism and rehydrated in a water-saturated atmosphere. Spectra were then recorded during drying over an approx. three minute period. (A) the 1800- 1000 cm^{-1} 'fingerprint' region; (B) the 3800-2800 cm^{-1} region. Rehydrated spectra are shown in blue and dried spectra in red.*

4.10.2 Sulphate

Sulphates are a product of amino acid metabolism in the body, particularly that of sulphur-containing methionine and cysteine. Urinary levels of sulphates are therefore influenced by dietary intake of protein [131]. The main absorbance bands of sulphate occur in the 1200 - 950 cm^{-1} region, so the dehydration spectra shown in Figure 4.16 have been plotted to include the whole of this region.

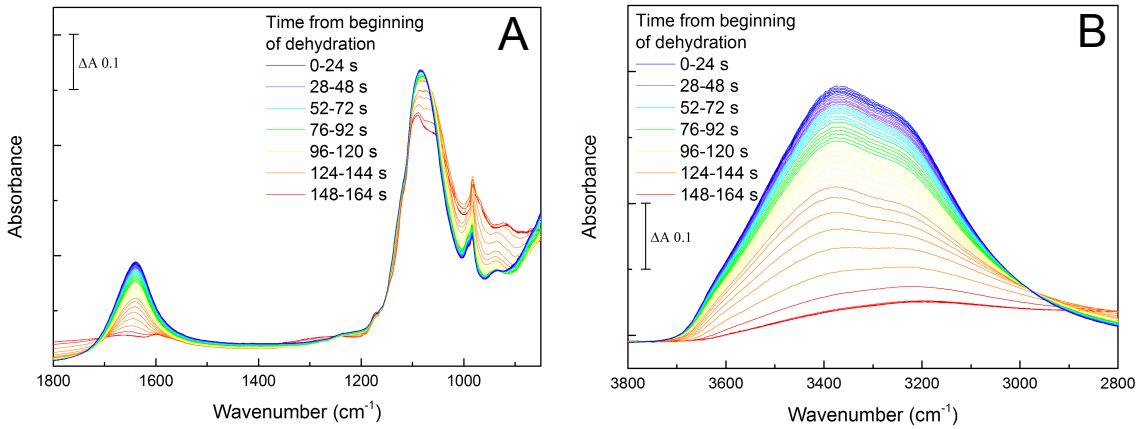


Figure 4.16: Serial spectra of a rehydrated potassium sulphate sample during drying. *3 µL of a 100 mM solution of potassium sulphate (adjusted to pH 7 with potassium hydroxide) was firstly dried on the prism and rehydrated in a water-saturated atmosphere. Spectra were then recorded during drying over an approx. three minute period. (A) the 1800- 1000 cm^{-1} 'fingerprint' region; (B) the 3800-2800 cm^{-1} region. Rehydrated spectra are shown in blue and dried spectra in red.*

4.10.3 Phosphate

Phosphates are important in a number of bodily functions including in bone formation, energy production and cell signalling. The body's requirement for phosphate is met by a normal diet and excess phosphate in the blood is filtered by the kidneys and excreted in the urine [132].

Most of the changes observed in the dehydration spectra of potassium phosphate (Figure 4.17) were increases in intensity of phosphate bands as the water was removed from the sample along with the appearance of small but sharp shoulders between 1200 and 1100 cm^{-1}. There was a small downshift of the band centred around 1080 cm^{-1} in the final stages of dehydration. Unusually, the absorbance intensity in the 3800-2800 cm^{-1} region increased as the sample dried (Figure 4.17 B) suggesting that phosphate contributes to the absorbance in this region.

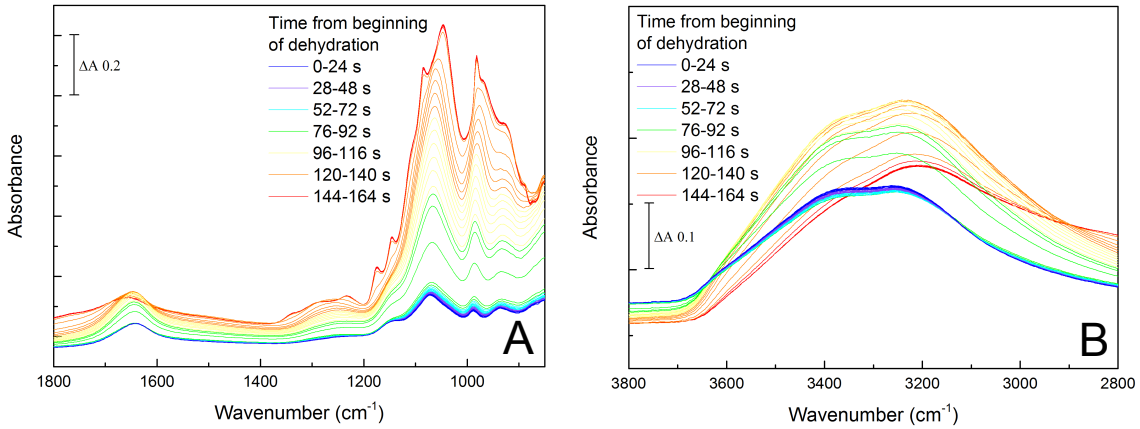


Figure 4.17: Serial spectra of a rehydrated potassium phosphate sample during drying. *3 μL of a 50 mM solution of potassium phosphate (adjusted to pH 7 with potassium hydroxide) was firstly dried on the prism and rehydrated in a water-saturated atmosphere. Spectra were then recorded during drying over an approx. three minute period. (A) the 1800- 1000 cm^{-1} 'fingerprint' region; (B) the 3800-2800 cm^{-1} region. Rehydrated spectra are shown in blue and dried spectra in red.*

4.10.4 Glucose

In healthy individuals, any glucose that is filtered by the glomerulus is reabsorbed in the proximal tubule. Some conditions, such as diabetes mellitus where blood glucose levels are elevated, are associated with elevated urinary glucose (glycosuria) because the amount of glucose filtered into the urine exceeds the capacity of the proximal tubule to reabsorb it. Other conditions which affect blood glucose or the kidney function can also lead to glycosuria [108].

Glucose is highly soluble in water and this is expected to be reflected in its IR spectrum at different levels of hydration (Figure 4.18).

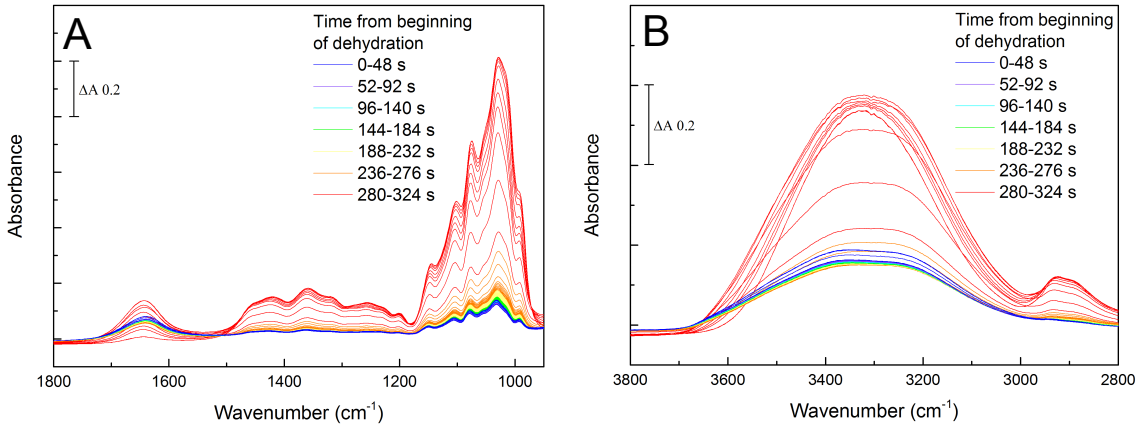


Figure 4.18: Serial spectra of a rehydrated glucose sample during drying. *3 μL of a 25 mM solution of glucose was firstly dried on the prism and rehydrated in a water-saturated atmosphere. Spectra were then recorded during drying over an approx. three minute period. (A) the 1800- 1000 cm^{-1} 'fingerprint' region; (B) the 3800-2800 cm^{-1} region. Rehydrated spectra are shown in blue and dried spectra in red.*

4.10.5 Cystine

Cystine is an amino acids hyperexcreted in to the urine of some individuals with a rare kidney stone disease called cystinuria. It is highly insoluble in water and urine at physiological pH and hence precipitates to form stones [133]. Unlike compounds such as urea and creatinine, it does not form hydrogen bonds or interact significantly with water molecules, therefore its IR spectrum, in terms of the position and shape of the absorbance bands, is not influenced by the extent of hydration (Figure 4.19). Since increasing hydration of a sample dilutes the concentration of other compounds, an intensity change in the cystine spectrum is expected with increased hydration as water molecules replace some of the cystine molecules in the active area of the prism.

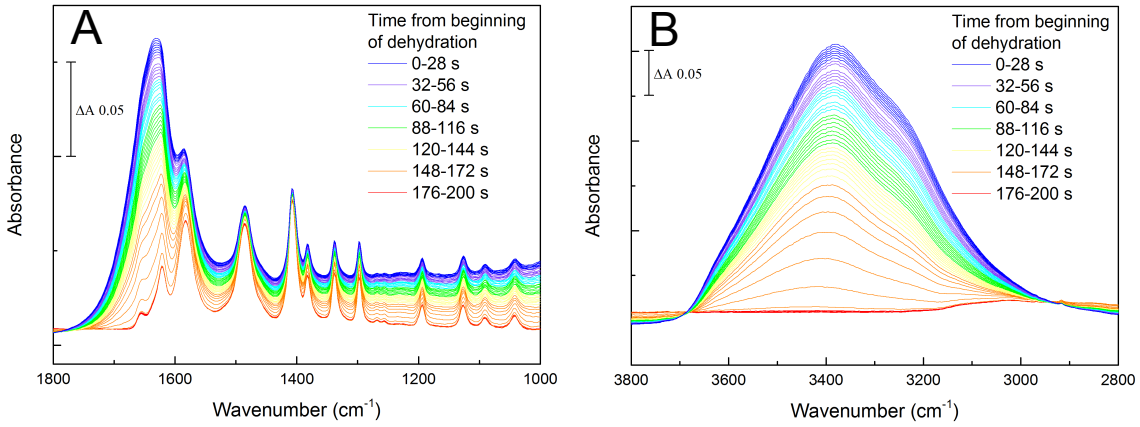


Figure 4.19: Serial spectra of a rehydrated cystine sample during drying. *3 μL of a 12 mM suspension of cystine was firstly dried on the prism and rehydrated in a water-saturated atmosphere. Spectra were then recorded during drying over an approx. three minute period. (A) the 1800- 1000 cm^{-1} 'fingerprint' region; (B) the 3800-2800 cm^{-1} region. Rehydrated spectra are shown in blue and dried spectra in red.*

4.11 Discussion and conclusions

In the context of IR analyses of urinary samples, at least four IR-distinct forms of urea should be considered, with their relative amounts dependent on hydration level. These correspond to the rehydrated form (equivalent to the solution form), a dried form and two additional partially hydrated forms. These different forms are likely to correspond to specific urea-urea and urea-water structures, the likes of which have been previously identified in the literature. In support of this model, Gaussian modelling of such urea structures and their predicted IR spectra suggests that the absorbance frequency of the υ_{as}(CN) band of urea downshifts from a fully hydrated to a fully dried state before upshifting as an increasing number of urea-urea interactions are formed.

A selection of the modelled urea-water and urea-urea structures are shown in Figure 4.20. Based on structures A-C in Figure 4.20 it appears that the hydration starts from the carbonyl group, rather than the amino terminal as suggested by Nandel *et al.* (1998). [121] However, the urea-water and urea dimer structures are consistent with those predicted by Lee *et al.* (1995) [119] and with the notion that the primary hydration sphere contains four water molecules [119,121]. The five-urea structure shown in Figure 4.20 D shows that when there are relatively few urea molecules the environment of the υ_{as}(CN) band differs for each, resulting in differences in their vibrational frequency.

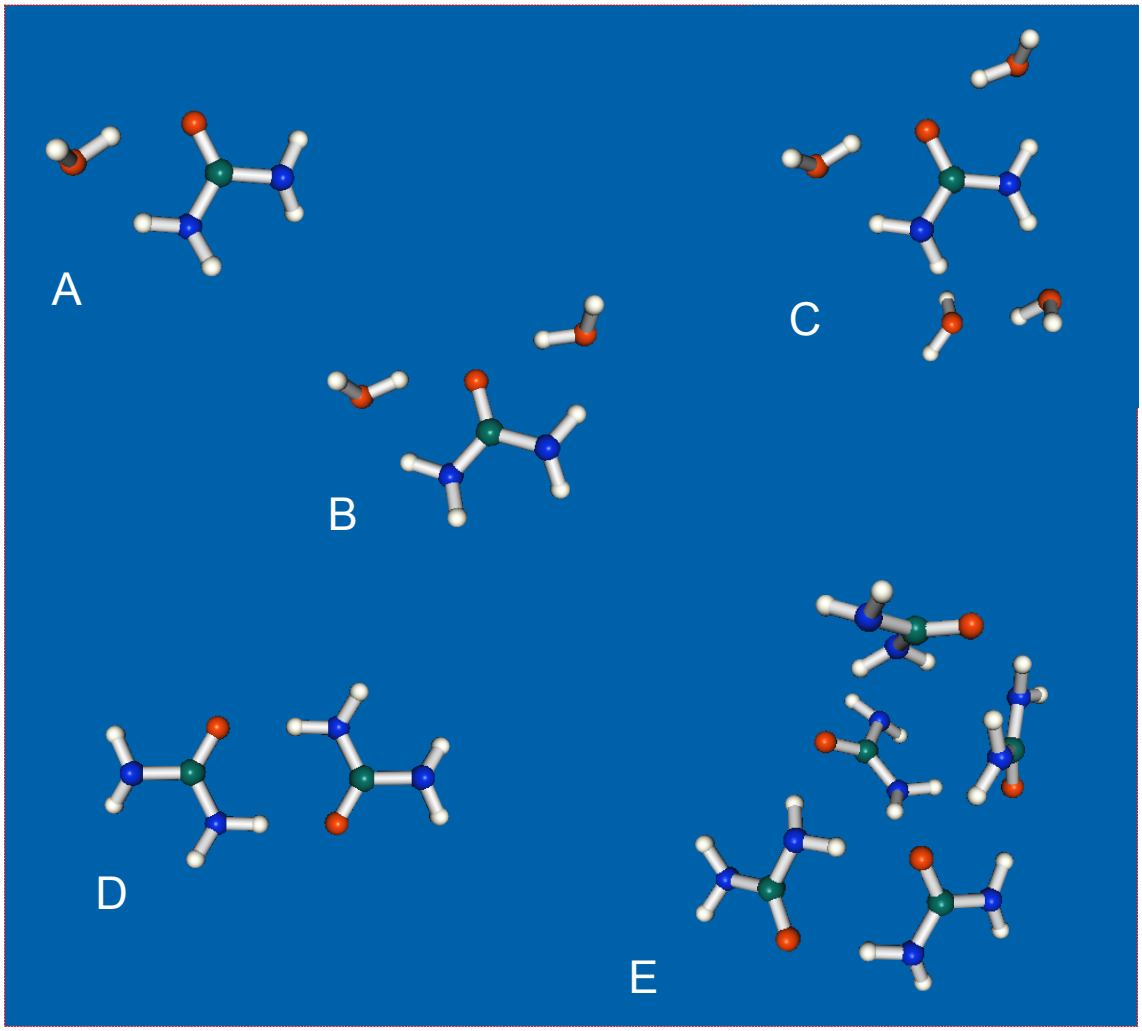


Figure 4.20: Modelled urea-urea and urea-water structures. *(A) Urea with a single water molecular. (B) Urea with two water molecules. (C) Urea with four water molecules. (D) Two urea molecules. (E) Five urea molecules.*

Multiple hydration states can also exist for other materials, as demonstrated here with creatinine. The (υ(C=N) (30) υ(CN) (27) δ(NCH) (24)) absorbance band of creatinine in dried and rehydrated/wet spectra can each be described by single components. In partially rehydrated creatinine spectra a combination of one of these components plus a single intermediate centred at 1501 cm^{-1} can be used to adequately fit the data.

Spectra of dried urine samples show that the partially hydrated states of urea tend to become trapped in the dense matrix of dried urinary components, with a predominance of the lowest wavenumber (1443 cm^{-1}) form, preventing direct urea-urea interactions. The retention of such partially hydrated forms with altered IR signatures complicates the analysis of infrared spectra of dried samples such as urine, and it is important that they are recognised in diagnostic analyses of dried biological tissues and fluids. However, the limited number of distinct forms can easily be accommodated in de-

composition procedures for quantitative analyses, as shown here with the contributions of different states of urea in dried urine samples.

Urea and creatinine in urine have been quantitated by other groups using FTIR spectroscopy [49, 81], but samples were analysed wet and so the problems of hydration would not have been encountered. Shaw *et al.* (2000) [91] discussed the quantitation of urea, creatinine and protein using FTIR spectroscopy of dried urine films. PLS regression analysis was applied to the 1400 - 1800 cm^{-1} region for creatinine quantitation and the 3100-3550 cm^{-1} region for urea. Both of these regions have been shown to be influenced by the extent of hydration of the sample, but control of hydration was not mentioned and no spectra were presented in the report for critical analysis. Similarly, dried mouse urine was analysed for urea by Furlan *et al.* (2004) [134] without mention of controlled hydration or presentation of spectra. Other groups have used FTIR spectroscopy to analyse dried blood products for urea, creatinine and other components without considering hydration effects [73, 74]. However, Fabian *et al.* (2005) [76] noticed that peak intensities in spectra of dried serum films were influenced by relative humidity and acknowledged that environmental conditions should be carefully controlled during analysis. Ollesch *et al.* (2013) [77] also discussed how reproducibility of dried serum and plasma sample spectra was adversely affected by inhomogenous dried films. The 'coffee-ring' drying patterns described probably resulted in varying degrees of hydration across the sample. They also demonstrated how bands can shift and change shape between wet and dried spectra of serum samples, with particular focus on amide I and II absorbance bands. They were able to improve reproducibility by using a robotic dispensing system that could generate dried sample spots of homogenous thickness. Control of environmental humidity combined with automated sampling could offer an alternative way of reducing the influence of the hydration effect and reducing the need for computational compensation post-measurement.

Chapter 5

Analyses of urine FTIR spectra for disease diagnosis.

Author contributions: the first section of this chapter has been prepared for publication. All data and figures in the publication were measured and prepared by Katherine Oliver. The paper was written by Katherine Oliver with contributions from Amandine Maréchal, Robert J. Unwin and Peter R. Rich. Shabbir H. Moochhala offered clinical guidance. Annalisa Vilasi performed preliminary work upon which the project was extended. Additional data has been included.

5.1 Case study: cystinuria

5.1.1 Introduction

Cystinuria is an autosomal inherited aminoaciduria caused by mutations in one or both of the subunits of the $b^{0,+}$ amino acid transport system. It results in a failure to reabsorb freely filtered cystine in the proximal tubule of the kidney, causing increased urinary excretion of cystine and the dibasic amino acids lysine, arginine and ornithine [133,135]. Cystine is a dimer of cysteine formed by oxidation of their sulphydryl groups to form a disulphide bond. Cystine is poorly soluble in water in the physiological pH range [136]. Its solubility limit in urine is ~ 1 mM [133,137] and it forms a microcrystalline precipitate above this concentration. Insoluble cystine can form stones that may cause obstruction of the kidneys and lead to irreversible damage and loss of function.

Cystinuria has a global prevalence of around 1 in 7000, ranging from 1 in 2000 to 1 in 100 000, depending on the population. Cystine stones account for only around 1-2 % of all renal stones in adults and 6-8 % in children. The first incidence of stone formation may occur at any age but over half of patients develop their first stone before the age of 20 [104,133].

Cystine levels in cystinuric patients typically range from 1-2 mM, but concentrations over 4 mM have been reported [138]. Patients with urinary cystine concentrations over 1 mM have a significantly increased risk of stone formation. Many patients respond to simple treatments in the form of increased fluid intake, especially overnight, urinary alkalinisation, an alkaline ash or low methionine (animal protein) diet, and dietary salt restriction. Dithiol drugs can also decrease the urinary cystine concentration by disulphide bond reduction to form more soluble drug-cysteine complexes. However, these measures do not treat the underlying cause of the disease and are most effective when used prophylactically [133,135]. Once stone formation has occurred, shock-wave lithotripsy or surgery is usually required.

Cystinuria is usually diagnosed by measuring urinary cystine. Early diagnosis

and preventive maintenance therapy are essential and both require the ability to detect and quantitate cystine quickly, easily and repeatedly. Qualitative colorimetric screening tests that can provide rapid results are available, but these have low specificity. For example, the sodium-nitroprusside test (Brand's technique) can produce a positive result in a few minutes. However, the use of cyanide has associated hazards, and the reaction is subject to interference from ketones, homocystine from the diet, ascorbic acid and acidic conditions, which can all give false positive results [133, 139]. The Shinohara and Padis method [140] is an alternative colorimetric method for estimating urinary cystine concentration by measuring the concentration of SH groups, but also suffers from interference. Derivatisation followed by ion exchange chromatography is the current clinical gold standard for amino acid analyses, including cystine. However, such chromatographic techniques are time-consuming and expensive in personnel and equipment [104]. As a result, they are not generally available at the point of care. Furthermore, before analysis by ion exchange chromatography urine samples must be treated to increase the solubility of cystine as this method is only able to measure soluble cystine. Generally, this is achieved by raising or lowering the pH of the sample but the solubility of cystine does not increase dramatically except at the very extremes of pH meaning ion exchange chromatography may underestimate the total cystine load in samples with a very high concentration of cystine. Alternatively, an individual's urinary cystine solubility threshold can be quantitated by a cystine capacity assay [141]. This involves incubating a known amount of solid cystine with a 24 hour patient urine sample. The sample is spun down and the insoluble fraction weighed to give a measure of cystine saturation in the urine. However, this, again, is a labour-intensive procedure that is not widely available. Hence, a faster, simpler and more cost-effective routine method for quantitating urinary cystine would be of significant clinical value.

ATR-FTIR spectroscopy is an ideal candidate for an alternative method of detecting and quantitating cystine in urine because it is quick, specific and has the potential sensitivity required for detection of the levels of cystine found in cystinuric urine. Cystine has a unique IR spectrum with a number of distinct bands. Attempts have already been made to detect cystine in kidney stones. Marickar et al. (2009) [142] used scanning electron microscopy-elemental distribution analysis in combination with FTIR spectroscopy in order to identify cystine in 30 renal stone samples. They compared their transmission spectra to a cystine standard spectrum and standard spectra of other stone-forming compounds in order to correlate sample peaks with those of the known compounds. Peaks quoted as being characteristic of cystine include 3026, 1618, 1485 and 846

cm^{-1}. Cystine stones were also detected by Paluszkiewicz and co-workers (1988) [53] in a study of Polish populations. However, FTIR spectroscopy has not been applied to the detection and quantitation of cystine in fresh urine samples, which is the simplest diagnostic test. In this report the potential of ATR-FTIR spectroscopy as a rapid, reagent-free method for detecting and quantitating insoluble cystine in urine has been explored.

5.1.2 Quantitation of urea and creatinine in urine by ATR-FTIR spectroscopy

Since patient fluid intake before urine collection cannot be easily controlled, and because glomerular filtration rate (GFR) varies and can affect solute excretion, urinary creatinine concentration is commonly used clinically to normalise concentrations of other compounds in urine. Urinary urea can also be an indicator of urine dilution. These can be determined from IR spectra of whole urine by various methods [49,81,91]. In this study, urea and creatinine were determined by deconvolution of the 1510-1445 cm^{-1} region of whole urine spectra, which is dominated by absorbance bands of urea and creatinine.

The 'Peak Analyzer' function of the Origin 8.5 software was used to find optimal parameters for component fitting of spectra of pure solutions of urea and creatinine in double distilled water. A spectrum of a pure solution of urea after subtraction of contributions of bulk water and water vapour was best fitted with a single Gaussian component centred at 1468 cm^{-1} with full width at half maximum height (FWHM) of 33 cm^{-1}. A pure spectrum of creatinine treated in the same way could be fitted with a single Lorentzian component centred at 1492 cm $^{-1}$ with FWHM of 24 cm^{-1}. Unrestricted linear baselines were used in both instances. A solution of 100 mM urea plus 10 mM creatinine was prepared as a simplified model for human urine to determine whether the same parameters could be used to a fit and simultaneously quantitate urea and creatinine in a mixture. The previously determined parameters for fitting urea and creatinine individually were used as starting parameters but were allowed to vary during the fitting of the mixed solution spectrum. It was found that the peak centres and lineshapes remained unchanged for fitting urea and creatinine in the mixed solution spectra but the FWHM of the urea component was reduced to 32 cm $^{-1}$ for optimal fitting. The final parameters for fitting urea and creatinine in mixed spectra are defined in Table 5.1.

Table 5.1: Final parameters for quantitative component fitting of urea and creatinine in mixed solution spectra.

Compound	**Vibrational assignment**	**Component centre (cm^{-1})**	**Function type**	**FWHM (cm^{-1})**
Urea	v_{as}(CN)	1468	Gaussian	32
Creatinine	(v(C=N) (30) v(CN) (27) δ(NCH) (24))	1492	Lorentzian	24

To validate the component fitting the parameters in Table 5.1 were used to fit a series of spectra of urea and creatinine mixtures at known concentrations. The integrals of the fitted components of the calibration sample (100 mM urea plus 10 mM creatinine) were used to estimated the concentrations of the test samples. The results are shown in Table 5.2. The method is generally accurate for mixtures of urea and creatinine but is less accurate when the mixture contains a very low concentration of urea or creatinine. By fitting to spectra of the seven prepared solutions of 100 mM urea plus 10 mM creatinine the average predicted concentrations of urea and creatinine were 100 mM (± 1 mM standard deviation) and 10.6 mM (±0.5 mM standard deviation) respectively.

Table 5.2: Validation of urea and creatinine component fitting.

Actual [urea] (mM)	Actual [creatinine] (mM)	Predicted [urea] (mM)	Predicted [creatinine] (mM)
200	30	200	30
200	30	203	30
100	15	101	15
100	15	103	15
100	10	100	11
100	10	101	10
100	10	100	10
100	10	100	11
100	10	99	11
100	10	99	10
100	10	100	10
100	0	95	1
50	0	46	0
0	10	6	10
0	5	1	5
0	2	1	2

The component fitting method is illustrated in Figure 5.1, which shows the ATR-FTIR spectra of pure urea and creatinine and a mixture in aqueous solution (Figure 5.1 A, B & C), along with the component fit of the 100 mM urea plus 10 mM creatinine mixture in Figure 5.1 D.

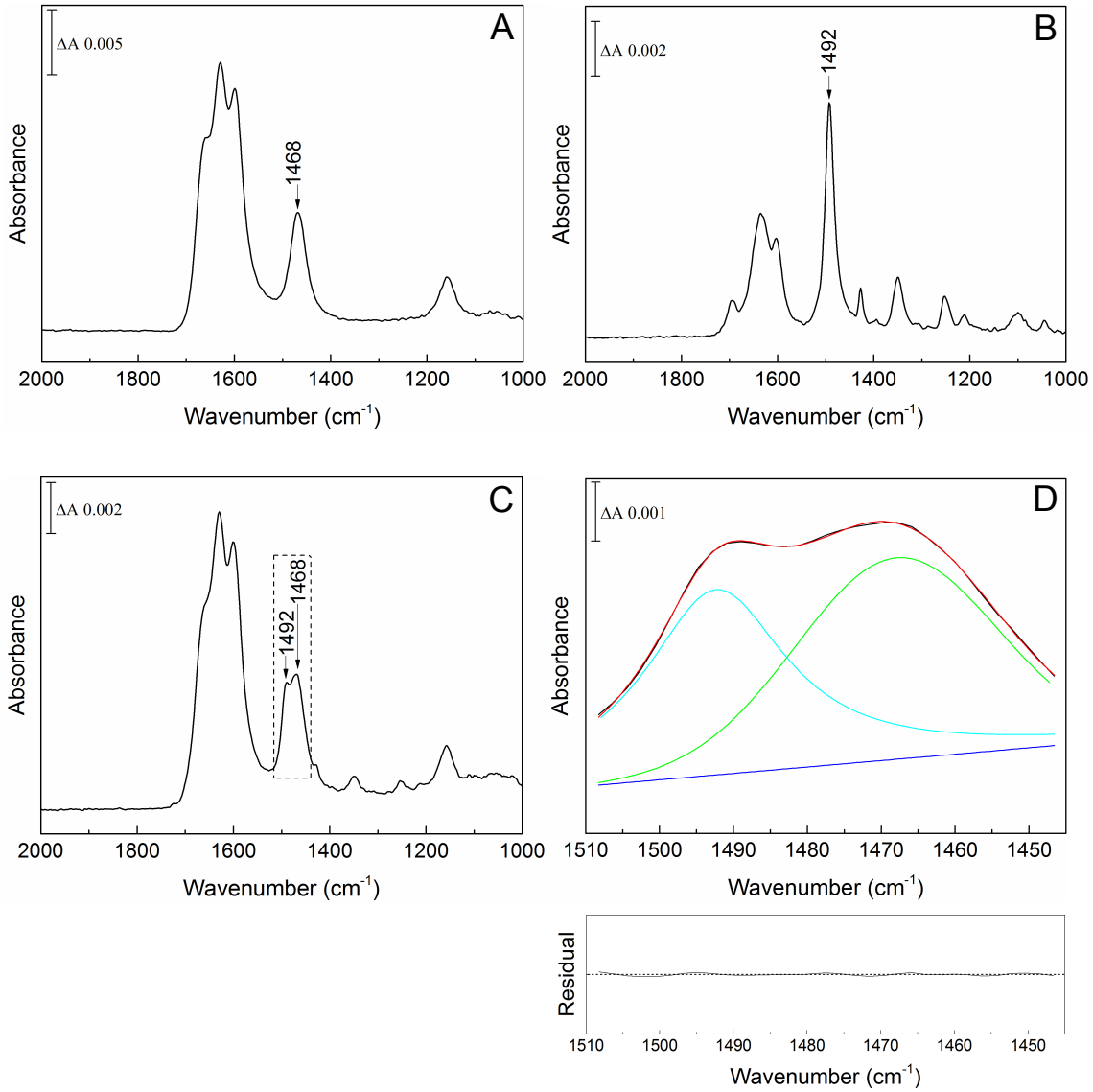


Figure 5.1: ATR-FTIR spectra of urea and creatinine in solution and deconvolution of mixtures. *Spectra of (A) 200 mM urea, (B) 30 mM creatinine and (C) a 100 mM urea/10 mM creatinine mixture in water are shown after subtraction of contributions of water solvent. In (D) data of (C) between 1510 and 1445 cm^{-1} (dashed box) were simulated with a combination of peaks from urea (Gaussian peak at 1468 cm^{-1} with FWHM of 32 cm^{-1}; shown in green) and creatinine (Lorentzian peak at 1492 cm^{-1} with FWHM of 24 cm^{-1}; shown in cyan):- recorded data (black); simulated curve (red); linear baseline (blue). The residual for the overall fit is shown underneath at the same scale as the component fit, with the centre dotted line representing zero.*

Calibration curves (Figure 5.2) were generated for urea and creatinine based on the integrals of fitted curves with fixed parameters as described in above Figure 5.1 and Table 5.1.

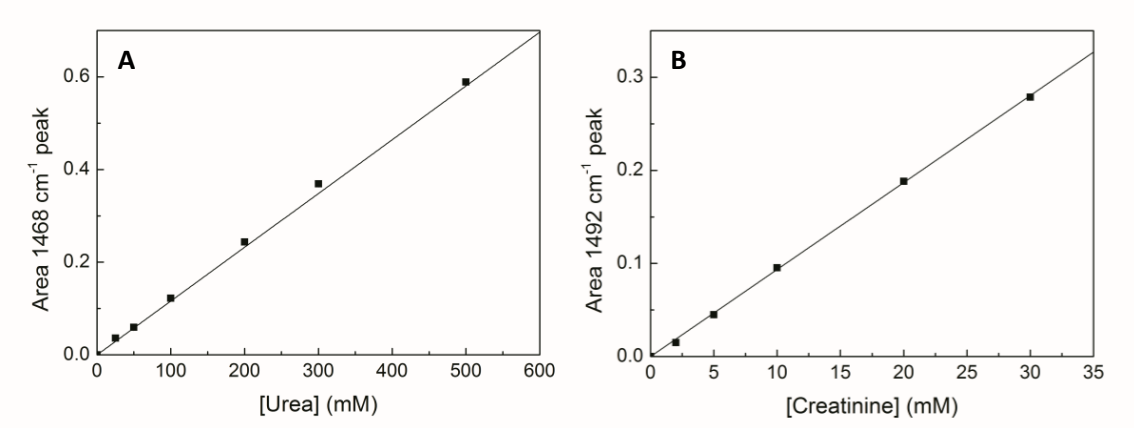


Figure 5.2: Calibration curves for urea and creatinine in water. *ATR-FTIR spectra of a range of aqueous urea and creatinine solutions were recorded. After subtraction of water contributions, curves were fitted between 1510 and 1445 cm^{-1}, with the parameters given in Figure 5.1. The areas of the fitted curved were plotted versus concentrations of (A) urea; (B) creatinine.*

Since urea and creatinine are the dominating components of the 1510-1445 cm^{-1} region of whole urine spectra (Figure 5.3 A), the same algorithm could be applied to determine their concentrations in urine samples. Figure 5.3 B shows an example of the fit to a typical ATR-FTIR spectrum of urine from a cystinuric patient. Figure 5.4 summarises the urea and creatinine concentrations determined by this method in urine samples from 22 cystinuric patients (labelled from P1 to P24) and 5 healthy controls. Urea concentrations were 28 – 443 mM, while creatinine concentrations were <1 – 20 mM, which are typical ranges. As expected, the concentrations of urea and creatinine tend to correlate and no pattern difference between cystinuric and healthy samples was evident.

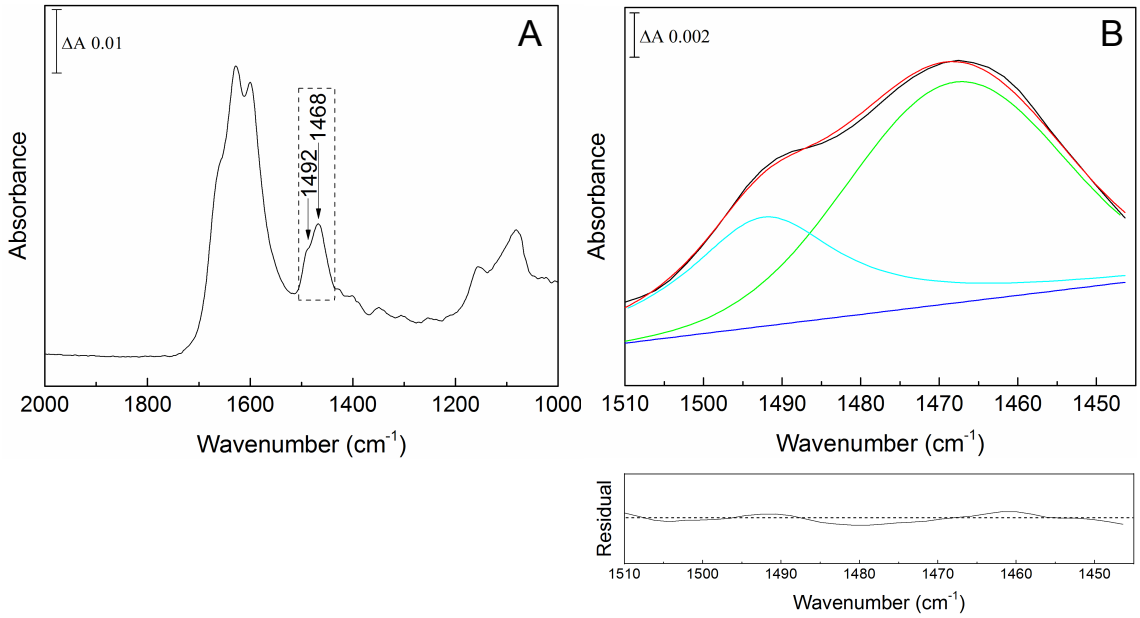


Figure 5.3: Deconvolution of urea and creatinine in whole urine spectra. *(A) A typical ATR-FTIR spectrum of a whole (undried) urine sample; (B) curve fitting of peaks of urea and creatinine in the 1510-1445 cm^{-1} region (dashed box in A) with fixed parameters as in Figure 5.1:- recorded data (black); simulated curve (red); linear baseline (blue); urea component (green); creatinine component (cyan). The residual for the overall fit is shown underneath at the same scale as the component fit, with the centre dotted line representing zero.*

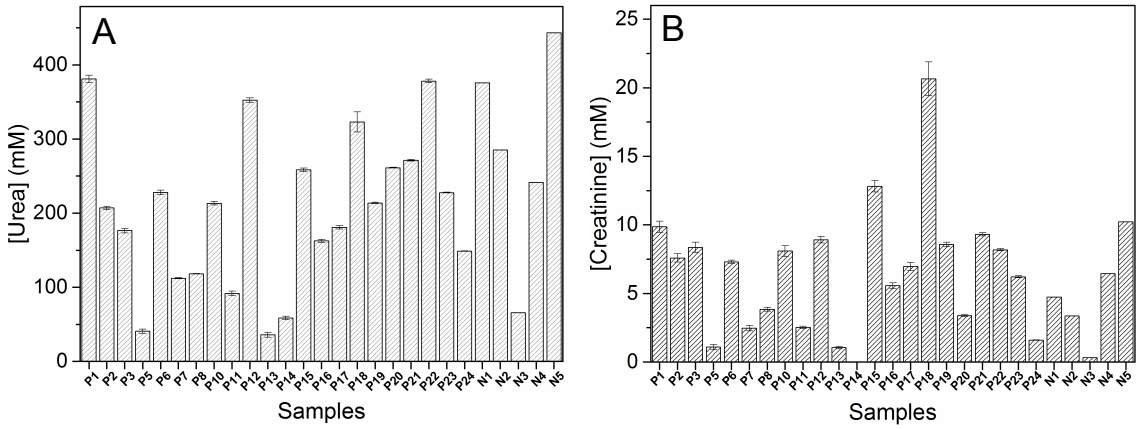


Figure 5.4: Urea and creatinine concentrations in urine samples. *(A) Urea and (B) creatinine concentrations in whole urine determined from ATR-FTIR spectra. Concentrations were deconvoluted as described in Figure 5.3 using the calibration plots of Figure 5.2. Pn: cystinuric patient; Nn: healthy control urine samples. Samples were measured in triplicate and error bars represent standard errors of mean.*

5.1.3 Comparison of ATR-FTIR and Jaffe reaction methods for creatinine quantitation

To validate the ATR-FTIR method for urinary creatinine quantitation, creatinine concentrations in the same urine samples were also determined using the Jaffe reaction since this is the most widely used clinical method [107, 143, 144]. A reasonable correlation was observed between the two methods (Figure 5.5 A) with a Pearson correlation coefficient of r=0.75 (r=0.83 if two outliers are removed), although the Jaffe method tended to predict slightly higher concentrations. A Bland-Altman analysis [99] was also performed to establish the similarity between the results generated by the two methods (Figure 5.5 B). The dashed line shows the mean bias and the dotted lines show 95% confidence limits (± 1.96 standard deviations from the mean). The mean bias was 1.4 mM, demonstrating that the Jaffe reaction tended to indicate higher creatinine concentrations. Two outliers had large discrepancies between the creatinine concentrations predicted by the Jaffe and ATR-FTIR methods; the origin of these discrepancies was not investigated further. The remaining data points all fell within the 95% confidence limits.

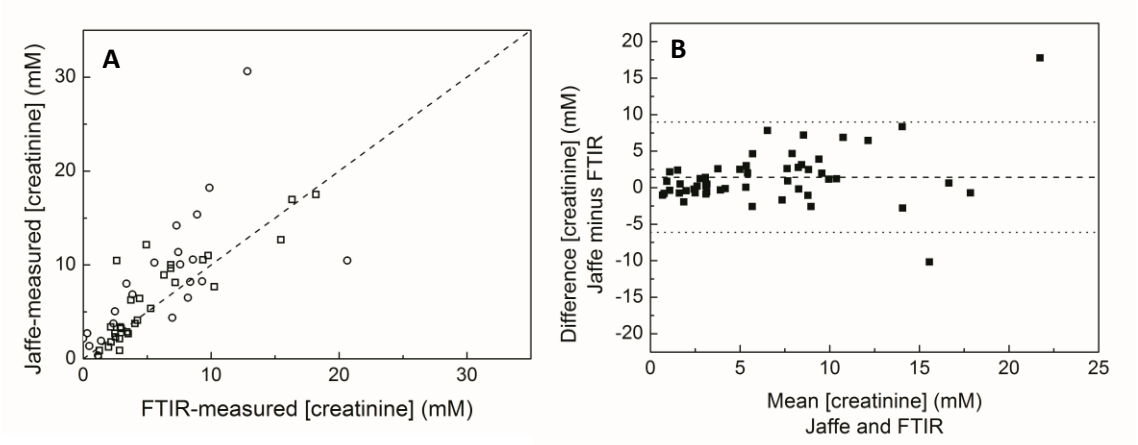


Figure 5.5: Comparison of ATR-FTIR spectroscopy and Jaffe methods of creatinine quantitation in whole urine samples. *(A) Circles represent those samples for which cystine content has also been measured. Squares represent additional urine samples. (B) Bland-Altman plot to assess the similarity between the Jaffe and FTIR methods for quantitating creatinine. The X-axis shows the mean of the creatinine measurements by the two methods whilst the Y-axis shows the differences of each method from the mean. Mean bias is shown by the dashed line with the 95% confidence limits shown by the dotted lines at mean bias ± 1.96 standard deviations.*

5.1.4 Development of measurement of insoluble compounds in urine

Cystine is poorly soluble in aqueous solutions between pH 2 and 12 [136]. Since only the first few microns of sample depth are probed by the evanescent wave of the IR beam in ATR mode, settling of any particulate cystine onto the prism surface will lead to unstable spectra. Hence, to avoid instabilities due to insoluble particle sedimentation, a quantitative protocol was developed in which samples were dried onto the prism surface (Materials and methods). However, initially, problems were encountered with poor reproducibility when quantitating cystine from dried sample spectra and several attempts were required to generate a reproducible cystine calibration curve.

The first variable tested in an attempt to improve reproducibility of cystine measurements was drying times. A 6.5 mM suspension of cystine was prepared and 5 µL aliquots were dried on to the prism by passing N_2gas over the sample at a controlled flow rate. The flow rate was varied to increase or decrease the drying time. Fast drying (over a 4 minute period) was achieved with a N_2flow rate of over 800 mL/min. Slow drying (over a 9 minute period) was achieved with a N_2 flow rate of 300 mL/min. The area of the 1303 - 1293 cm^{-1} cystine band in the final dried absorbance spectrum was calculated for each repeat and the mean and standard deviation for the slow and fast drying groups were calculated. Although the results (Figure 5.6) show an intensity difference between the two groups, the main conclusion is that slow drying of cystine suspension samples leads to better reproducibility as evidence by the smaller standard deviation of the slow drying group compared to the fast drying group.

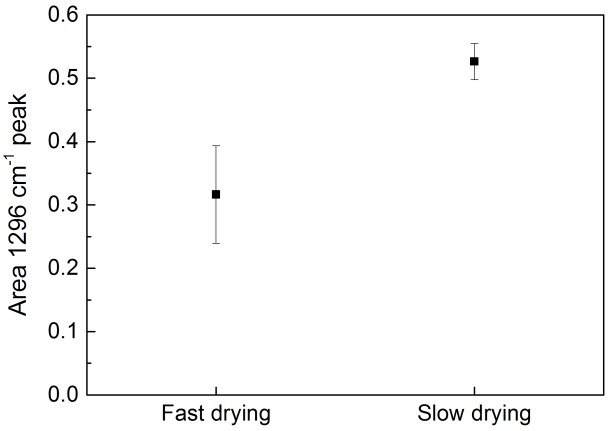


Figure 5.6: The effect of drying time on reproducibility of cystine spectra. *Reproducibility was assessed by calculating the mean area and standard deviation between 1303 and 1293 cm^{-1} of the spectra of cystine dried at a fast (4 mins or 800+ mL/min N_2 flow rate) or slow (9 mins or 300 mL/min N_2 flow rate) flow rate. Four repeats were performed for each group.*

It was then observed that the intensity of the 1296 cm^{-1} band of dried cystine varied between cystine stock solutions. Although, all stock solutions were prepared following the same protocol, at consistent volume, concentration and pH, it might be that cystine precipitation as the stock solution was neutralised from acidic pH is very sensitive to subtle differences, such as in the speed of neutralisation, that result in differences in particle formation. This could, in turn, influence the amount of cystine that is picked up in any 5 µL aliquot. To counter this, 1 mL aliquots of the same stock suspensions were homogenised by hand for 5 minutes to break up any large cystine particles and ensure all cystine particles were approximately equal size. New spectra were recorded of dried 5 µL aliquots of these homogenised suspensions. This process significantly improved the reproducibility of the dried cystine measurements as can be seen in Figure 5.7 (columns 1 and 2), which compares the area and standard deviation of the 1296 cm^{-1} band in spectra of dried 5 µL aliquots of 20 mM unhomogenised and homogenised cystine suspensions. However, further investigation revealed that application of gentle pressure to the dried samples greatly increased the intensity of the 1296 cm^{-1} band (Figure 5.7, column 3). Applying pressure to the samples also improved the reproducibility of the measurements whether the sample was homogenised prior to drying or not. This suggested that the cystine layer can partition away from the prism surface as it dries.

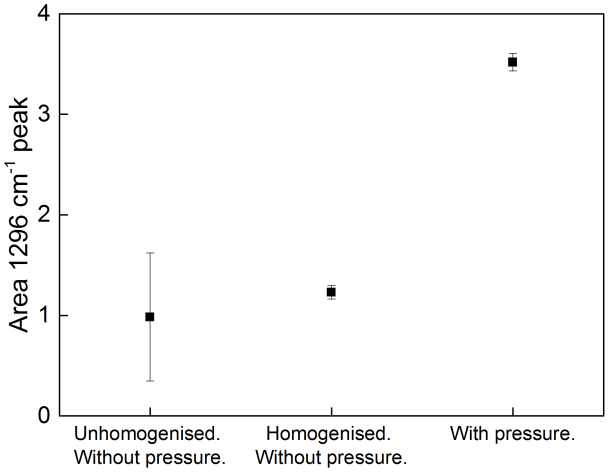


Figure 5.7: The effect of homogenisation and gentle pressure on the reproducibility of cystine spectra. *Reproducibility was assessed by calculating the mean area and standard deviation between 1303 and 1293 cm^{-1} of the spectra of dried cystine. In each case, 5 µL aliquots of 20 mM cystine suspensions were dried on to the prism by passing a gentle stream of N_2 across the sample at 300 mL/min. In the relevant cases, even pressure was applied to the dried sample using a standard press from SensIR.*

Based on these results, all subsequent IR measurements of cystine were per-

formed by drying cystine-containing samples on to the prism with a gentle stream of N_2 at 300 mL/min and gentle, even pressure was applied to the dried samples using a standard press from SensIR to ensure good contact between the sample and the prism surface.

5.1.5 Analysis of cystinuric urine samples

The structure of cystine is shown in Figure 5.8 and the ATR-FTIR spectrum of a dried L-cystine sample is shown in red in Figure 5.9 A. For reference, the dried solution and solid spectra of arginine, lysine and ornithine are also shown in Figure 5.10 as these amino acids are also hyperexcreted into urine in cystinuria. A calibration curve (Figure 5.11) for cystine was generated from heights of its 1296 cm^{-1} band, measured from second derivatives of absorbance spectra of a set of dried cystine suspensions with concentrations ranging from 0 to 3 mM at pH 7. The data were fitted with a linear function confirming that ATR-FTIR spectroscopy can be used to quantitate cystine in this concentration range, up to at least 3 mM. This calibration curve is likely to be applicable to urine samples at any physiological pH since the solubility of cystine only increases significantly when pH is below 2 or above 12 [136].

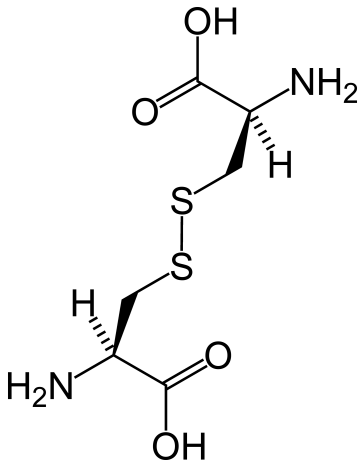


Figure 5.8: Structure of cystine.

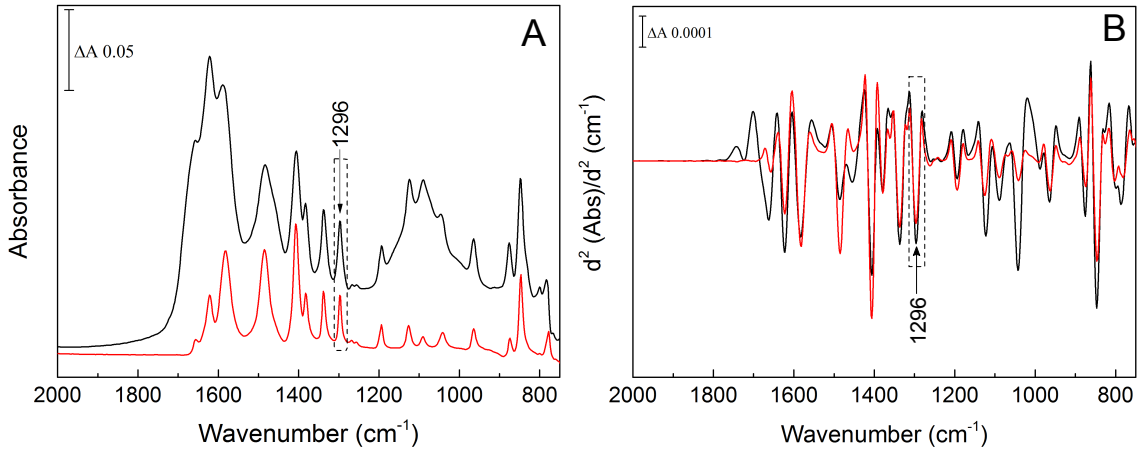


Figure 5.9: Absorbance spectra of cystine and insoluble cystinuric urine. *(A) Absorbance spectrum of dry cystine (red) and a dried insoluble fraction of a cystinuric urine sample (black). (B) Second derivative transformation of absorbance spectrum of dry cystine (red) and a dried insoluble fraction of a cystinuria urine sample (black).*

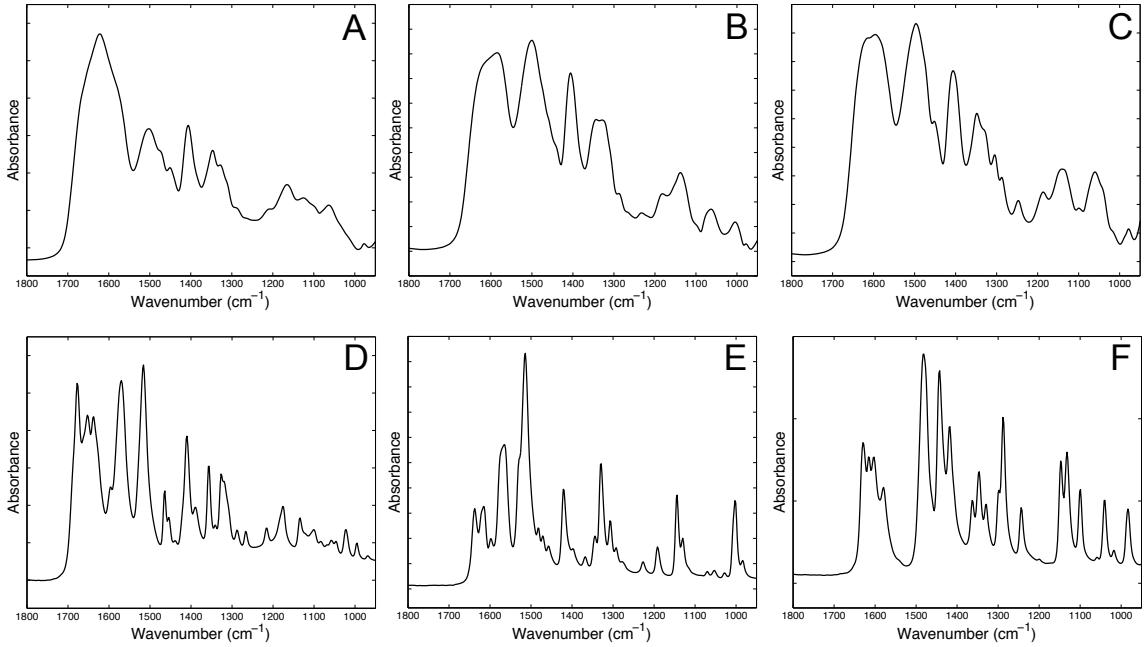


Figure 5.10: Absorbance spectra of other urinary amino acids associated with cystinuria. *(A) Spectrum of dried solution of L-arginine. (B) Spectrum of dried solution of L-lysine. (C) Spectrum of dried solution of L-ornithine. (D) Spectrum of solid L-arginine. (E) Spectrum of solid L-lysine. (F) Spectrum of solid L-ornithine. All solutions were adjusted to pH6 with phosphate buffer. The solid compounds were prepared with chloride as the counter ion, therefore the differences between the dried solution and solid spectra are likely to be a result of pH, or counter ion/buffer interaction.*

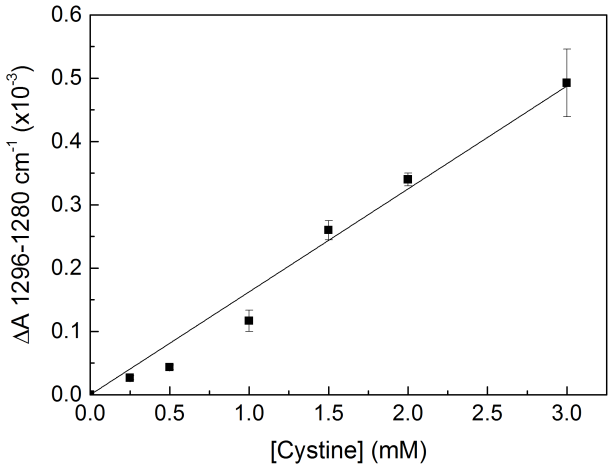


Figure 5.11: *Cystine calibration curve. Calibration curve for cystine dried from suspensions in water. To generate the calibration curve suspensions of 0 - 3 mM cystine in water at pH7 were prepared and 5 µL aliquots (homogenised by vortex mixing) were dried onto the prism. Pressure was applied on the sample before spectra were recorded. The height of 1296 cm^{-1} cystine band in the second derivative spectra were plotted against cystine concentration (intensity 1296 - 1280 cm^{-1}). A linear fit through the origin was calculated with no error weighting.*

Even in cystinuric urine, cystine levels will be low in comparison with many other urinary components and can be difficult to detect in whole urine spectra. Drying the urine sample on the ATR prism before analysis will increase the concentration of cystine in the active area of the ATR prism and, therefore, the intensities of its absorbance bands. However, absorbance bands of urea, creatinine and other dominating components that might mask some cystine bands will also be amplified. Since the clinically important cystine is the insoluble fraction, the urine samples were centrifuged to pellet the insoluble components, which could then be resuspended and dried onto the prism for quantitation. This removed contributions from soluble components and allowed accurate analysis of insoluble cystine from its characteristic band pattern (Figure 5.9, red), which could clearly be seen in the spectra of many of the dried, insoluble fractions of cystinuric urine samples (Figure 5.9, black). This demonstrates that urinary cystine can easily be detected by this method when it exceeds the critical solubility limit.

In brief, the insoluble material was first pelleted from 1 mL aliquots of each urine sample by centrifugation, hence removing any possible interference from soluble urinary components. Pellets were resuspended in 1 mL water. The sample was thoroughly vortexed before 5 µL were taken and dried onto the ATR prism surface and an FTIR spectrum was recorded. The cystine calibration curve (Figure 5.11) was used to estimate the concentration of insoluble cystine in the original urine samples. Twenty

two cystinuric (see Table 5.3 for summary of patient demographics) and 5 control urine samples (4 females aged 24-35 years, 1 male, unknown age) were analysed (Figure 5.12). Control urine samples were donated by healthy volunteers at the Royal Free Hospital, London.

Seventeen of the cystinuric samples had detectable insoluble cystine of up to 3.1 mM. Five cystinuric samples and all healthy controls had no detectable insoluble cystine, indicating that their urinary cystine concentrations were below the solubility limit of $\leq$1 mM.

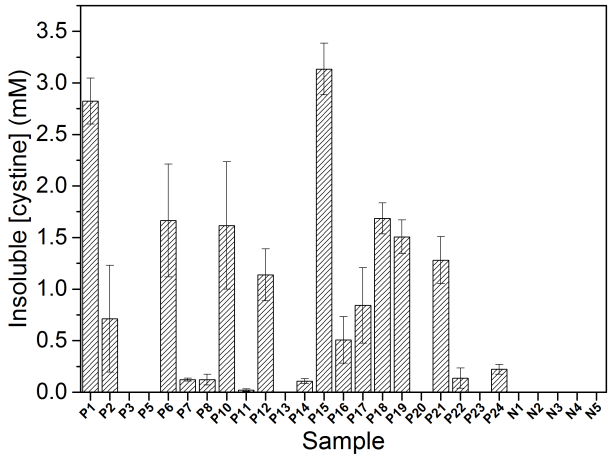


Figure 5.12: Levels of cystine measured in cystinuric (Pn) and control (Nn) urine samples by ATR-FTIR spectroscopy. *Patient urine samples were centrifuged and the pellets resuspended in water. Aliquots of the resuspended pellets were dried onto the ATR prism and pressure was applied before spectra were recorded. For cystine quantitation the spectra were converted to second derivative and the heights of the 1296 cm^{-1} bands were compared to the cystine calibration curve. Samples were measured in triplicate and error bars represent standard errors of mean.*

5.1.6 Comparison of FTIR and clinical methods of cystine quantitation

The cystine concentrations determined by ATR-FTIR spectroscopy were compared with the values obtained using the clinical gold standard method for urinary cystine estimation by ion exchange chromatography after derivatisation (performed externally at the Camelia Botnar Laboratories at Great Ormond Street Hospital) (Figure 5.13) [103]. Since the ATR-FTIR method measures only insoluble cystine, while the clinical method reports total cystine, in order to compare the two methods an additional 1 mM was added to each of the values obtained by ATR-FTIR (where there was a detectable level of insoluble cystine). Hence, samples for which ATR-FTIR detected no

insoluble cystine (5 cystinuric samples and 5 controls) may contain up to 1 mM soluble cystine, represented by points at 0.5 mM ± 0.5 mM in Figure 5.13 A. After offsetting the ATR-FTIR values by 1 mM to allow for this, there was a good correlation between the results obtained by the two methods (Pearson correlation coefficient r=0.915), as shown by the Bland-Altman plot (Figure 5.13 B).

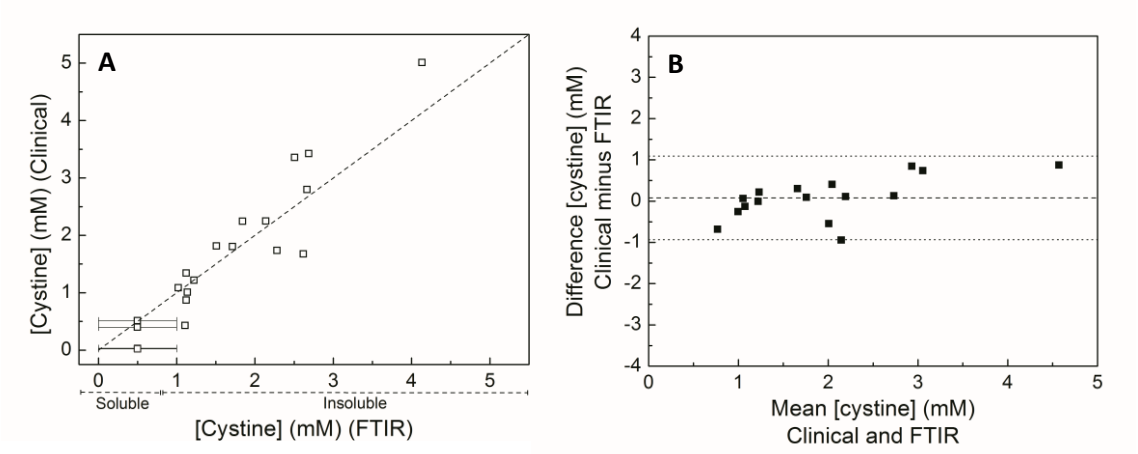


Figure 5.13: Comparison of cystine quantitation in urine by ion exchange chromatography and by ATR-FTIR spectroscopy. *(A) Comparison of cystine concentrations measured by ion exchange chromatography and ATR-FTIR spectroscopy. A 1 mM offset has been applied to the ATR-FTIR values for cystinuric samples to account for the additional soluble fraction that was discarded after centrifugation. Samples containing no insoluble cystine (5 cystinuric samples and 5 controls) are plotted as 0.5 ± 0.5 mM cystine to represent an unknown level of soluble cystine ranging from 0 up to 1 mM. (B) Bland-Altman plot to assess the similarity between ion exchange and FTIR methods (including 1 mM cystine offset) for quantitating cystine. Only samples containing measurable levels of insoluble cystine (determined by ATR-FTIR method) have been included. The X-axis shows the mean of the cystine measurements by the two methods whilst the Y-axis shows the differences from mean of each method. Mean bias is shown by the dashed line with the 95% confidence limits shown by the dotted lines at mean bias ± 1.96 standard deviations.*

5.2 Application of methods to other urinary disease biomarkers. Case study: APRT-deficiency

The methods described here for the detection and quantitation of insoluble cystine are likely to be adaptable for the detection of other diagnostic materials in urine, provided those materials have distinct absorbances in the mid-IR range and are present at sufficiently high concentrations. The present method developed for cystine is most suitable for measuring insoluble materials so most other kidney stone diseases are good

targets for further investigation. A short investigation into the development of the ATR-FTIR method for diagnosis of APRT-deficiency is described in this section as an example.

5.2.1 Introduction to APRT-deficiency

Adenine phosphoribosyltransferase deficiency (APRTd) is a rare, inherited disorder in which mutations in the gene encoding APRT lead to loss of function or production of the enzyme. APRT catalyses the conversion of adenine to adenosine monophosphate (AMP) but when the enzyme is deficient adenine is instead converted to 2,8-dihydroxyadenine (2,8-DHA), which is excreted into the urine. 2,8-DHA is poorly soluble in urine so can crystallise to form kidney stones. It is also toxic and can potentially lead to a decline in kidney function. Diagnosis of APRTd can be achieved by urine microscopy and identification of 2,8-DHA crystals by their distinctive round shape and reddish-brown colour. Biochemical stone analysis cannot always distinguish 2,8-DHA from uric acid, another stone-forming compound, but UV spectroscopy and x-ray crystallography can distinguish the two. Other diagnostic tests include measuring APRT activity and genetic testing [145–147]. These methods are generally either non-specific or relatively expensive and time-consuming. A sensitive and specific method for detection of 2,8-DHA in urine samples is not widely available. Here, a simple method for detection of 2,8-DHA in urine samples using ATR-FTIR spectroscopy is described.

5.2.2 Analysis of APRTd patient stone and urine samples

In order to generate a reference, an FTIR spectrum of pure synthesised 2,8-DHA (structure shown in Figure 5.14) was recorded (Figure 5.15, black). This was compared to a spectrum of a 2,8-DHA stone obtained from an APRTd patient (Figure 5.15, red). The comparison confirmed that the stone sample was formed of 2,8-DHA and was relatively pure. It also showed that the biological sample of 2,8-DHA that had formed a stone had a very similar spectral signature to a pure 2,8-DHA sample.

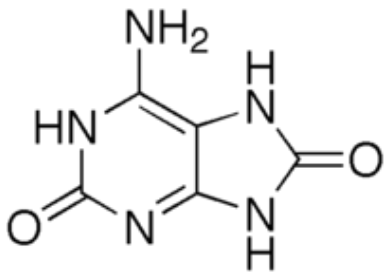


Figure 5.14: Structure of 2,8-dihydroxyadenine.

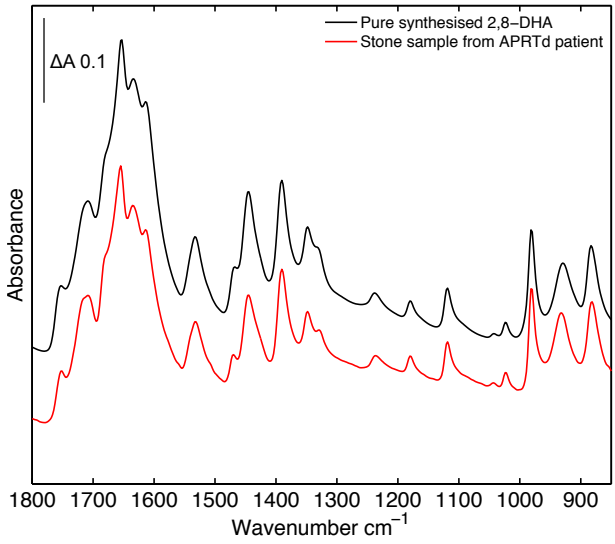


Figure 5.15: Reference absolute absorbance spectra of 2,8-DHA. *The black trace is a spectrum of dry, synthesised 2,8-DHA, whilst the red trace shows a stone sample from a patient diagnosed with APRTd.*

The second derivative of the absorbance spectrum of the synthesised 2,8-DHA sample was calculated to facilitate easier deconvolution of overlapping bands, as shown in Figure 5.16.

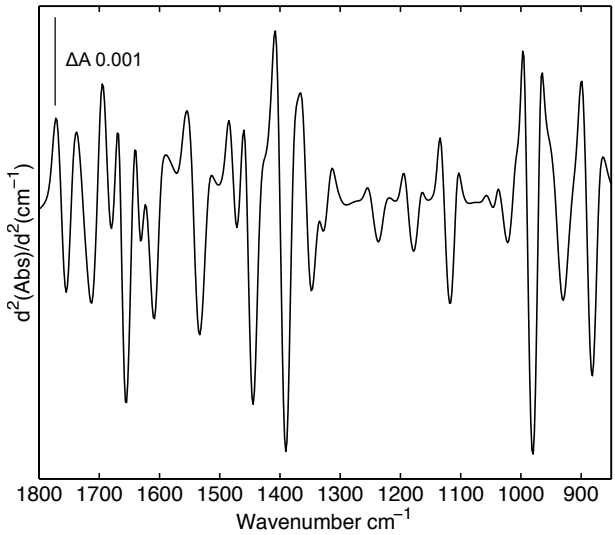


Figure 5.16: Second derivative of the absorbance spectrum of synthesised 2,8-DHA.

Urine samples (A-F) from patients diagnosed with APRT-deficiency (APRTd) were collected and their IR spectra recorded to determine whether it is possible to detect DHA in urine. Urine sample C was collected from a suspected APRTd patient. Urine samples D-F were collected from three members of a family of APRTd sufferers. The urine samples were centrifuged before analysis to remove most of the dominating soluble

components and dried, to remove the absorbance bands of liquid water and increase the intensities of all non-volatile component. Figure 5.17 A shows the absolute absorbance spectrum of the resuspended pellet derived from urine sample A after drying; Figure 5.17 B shows the second derivative of the same spectrum with the pure 2,8-DHA spectrum overlaid.

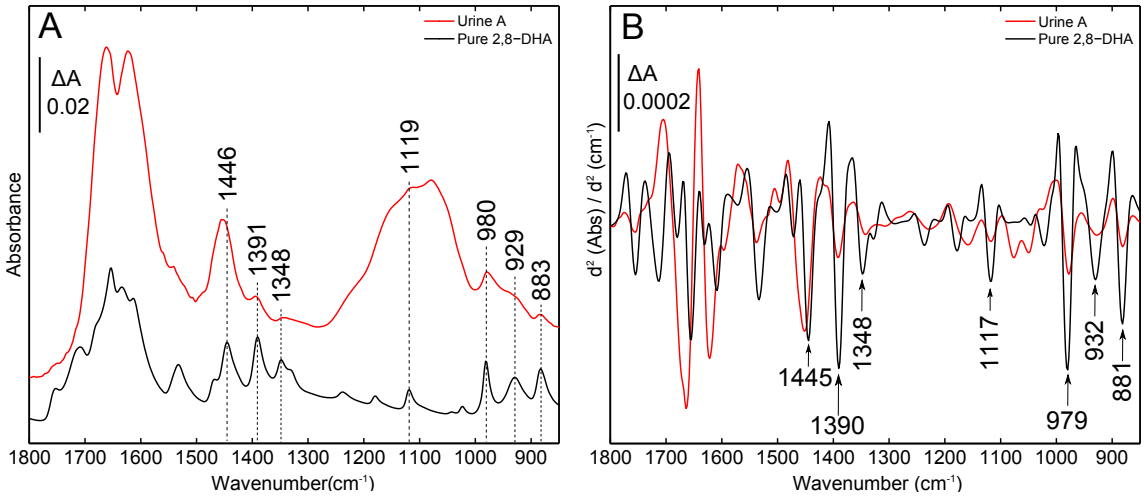


Figure 5.17: IR spectra of APRTd patient urine A. *Superimposed IR absorbance spectra of dried patient pelleted urine sample (red) and synthesised 2,8-DHA (black) as absolute (A) and second derivative (B).*

Several characteristic bands of 2,8-DHA are evident in the absolute urine spectrum, most clearly those at 883 cm^{-1}, 980 cm^{-1} and 1119 cm^{-1}. Since the second derivative transformation tends to sharpen bands (absorbance peaks become troughs in the transformation) and remove broad baseline effects additional bands related to DHA can be seen in the second derivative including those at 1390 cm^{-1}, 1117 cm^{-1}, 979 cm^{-1}, 932 cm^{-1} and 881 cm^{-1}.

Similar analyses of additional urine samples B-F, including three from members of the same family (samples D-F), did not show any detectable trace of 2,8-DHA, that is, there were insufficient spectral features similar to the IR spectra of 2,8-DHA present in the spectra of the urine samples. Figure 5.18 A, B & E-G shows the absolute spectra of the insoluble fractions of urine samples B-F, overlaid with the spectrum of pure 2,8-DHA; the equivalent spectra in second derivative are shown in Figure 5.18 C, D & H-J. There are one or two bands that correlate between the reference 2,8-DHA spectrum and each of the urine sample spectra, for example bands at 1246 cm^{-1} and 1117 cm^{-1}, however, other stronger bands, such as those at 881 cm^{-1}, 932 cm^{-1} and 979 cm^{-1}, are absent.

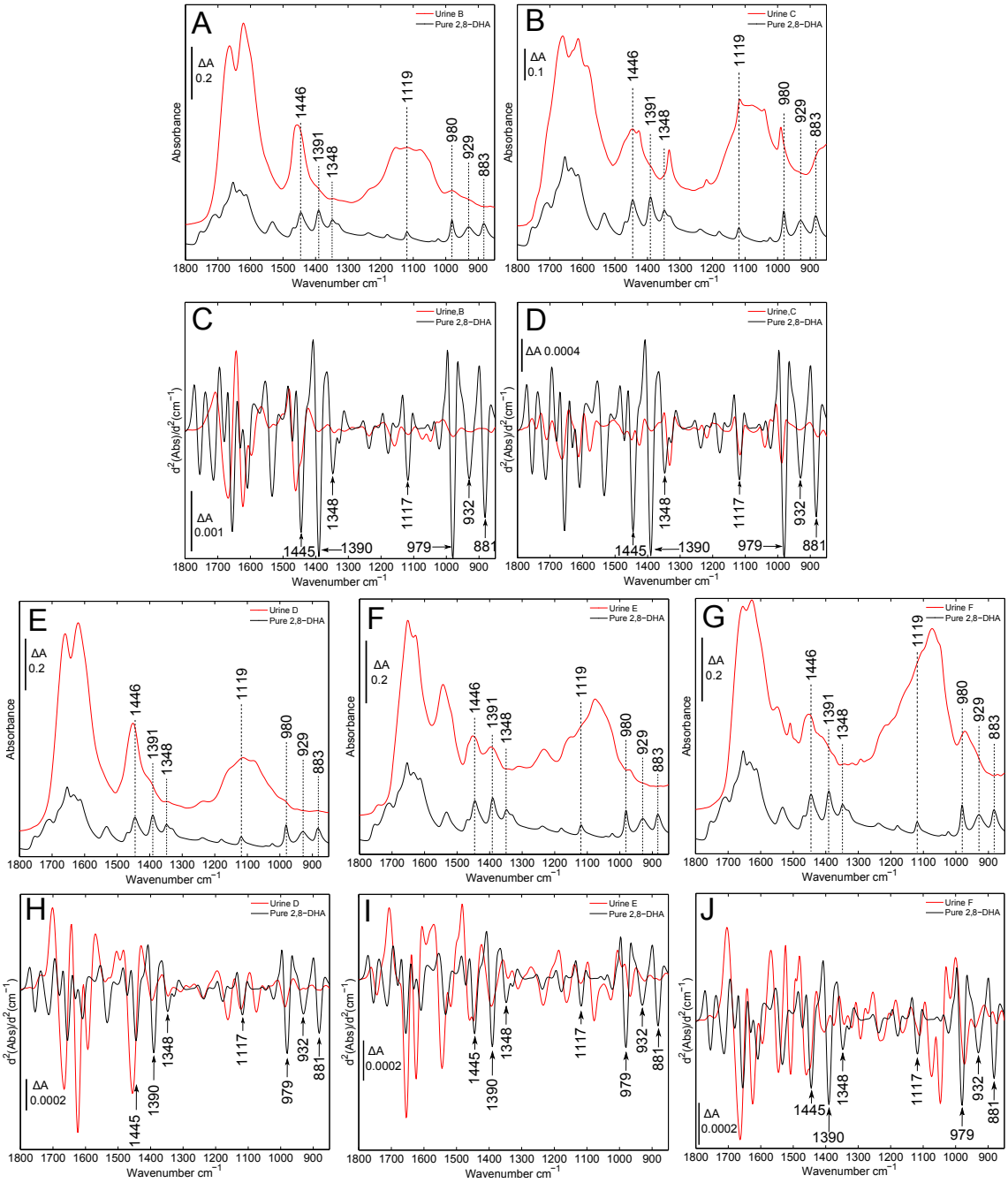


Figure 5.18: IR spectra of APRTd patient urine B-F. *Superimposed IR spectra of dried patient pelleted urine samples D-F (red) and synthesised 2,8-DHA (black). (A & C) Urine sample B. (B & D) Urine sample C. (E & H) Urine sample D. (F & I) Urine sample E. (G & J) Urine sample F. First and third rows show absolute spectra and second and fourth rows show the second derivative transformations of the absorbance spectra immediately above.*

Analyses of the supernatants of urine samples D-F also showed no indication of presence of 2,8-DHA (Figure 5.19)

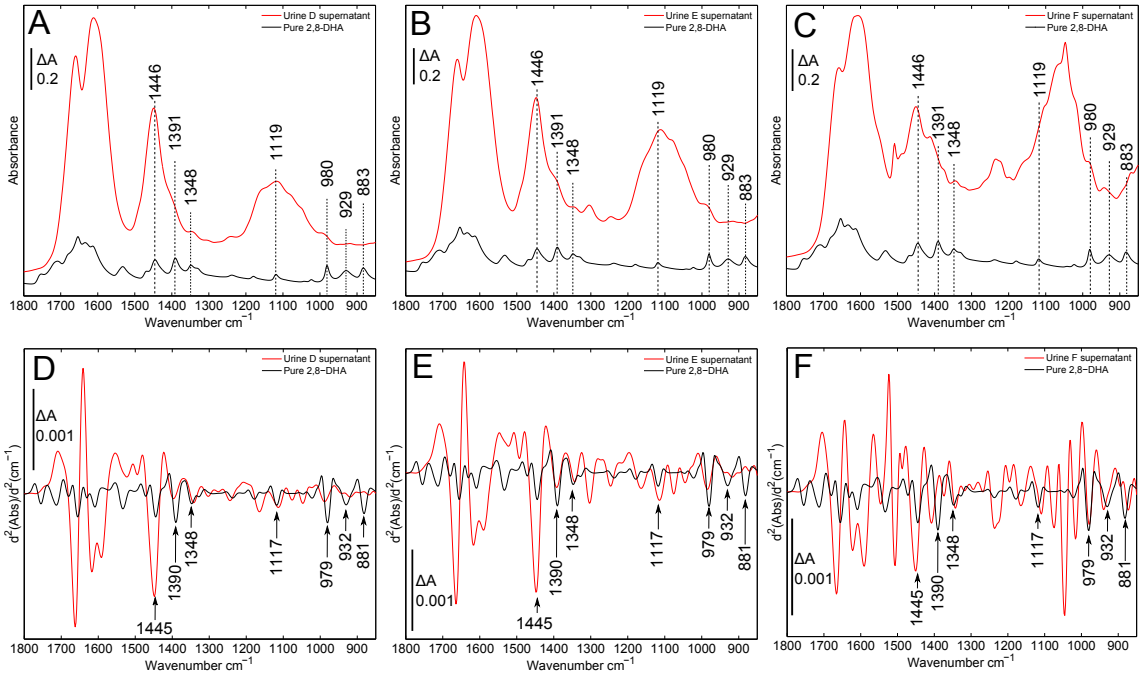


Figure 5.19: IR spectra of APRTd patient urine D-F supernatants. *Superimposed IR spectra of dried supernatant of patient urine samples D-F (red) and synthesised 2,8-DHA (black).Top row shows absolute spectra and bottom row shows the second derivative transformations of the absorbance spectra immediately above.*

In an attempt to determine the composition of the insoluble materials in urine samples B-F their spectra were compared to spectra of other stone-related compounds including oxalate, citrate, urate and uric acid (Figure 5.20). No clear correlations were observed.

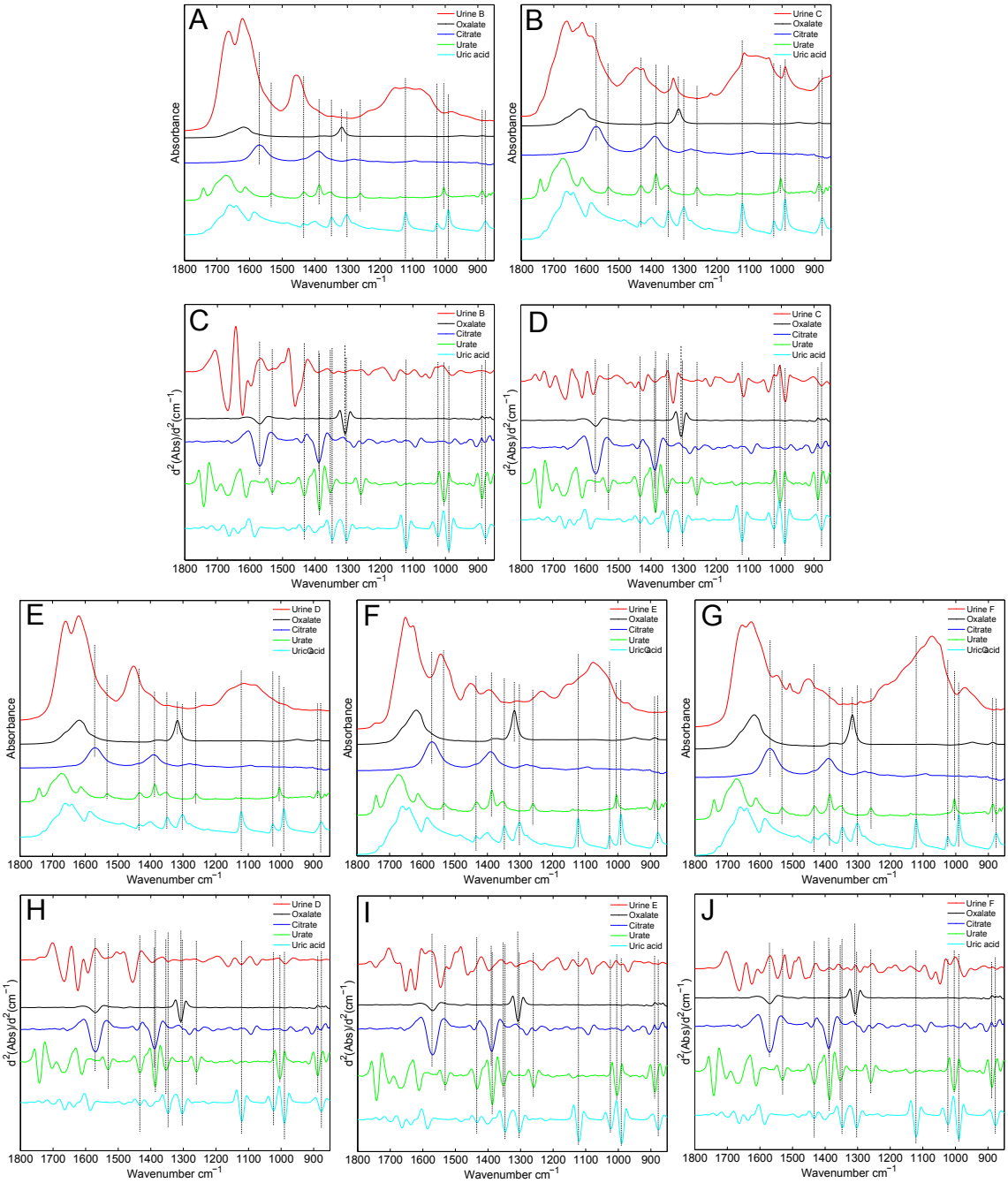


Figure 5.20: Comparison of APRTd patient urine D-F with model compounds. *Spectra of the insoluble fractions of urine samples D-F compared to known stone-forming compounds, oxalate (black), citrate (blue), urate (green) and uric acid (cyan). (A & C) Urine sample B. (B & D) Urine sample C. (E & H) Urine sample D. (F & I) Urine sample E. (G & J) Urine sample F. First and third rows show absolute spectra and second and fourth rows show the second derivative transformations of the absorbance spectra immediately above.*

In addition to human urine samples, urine samples from rat models of APRTd were also collected and provided by Dr. Richard Fish, UCL. These samples were treated in the same way as the human urine samples. The spectra of the insoluble fractions of

urine samples from two rats (rat X and rat Y) are shown in Figure 5.21.

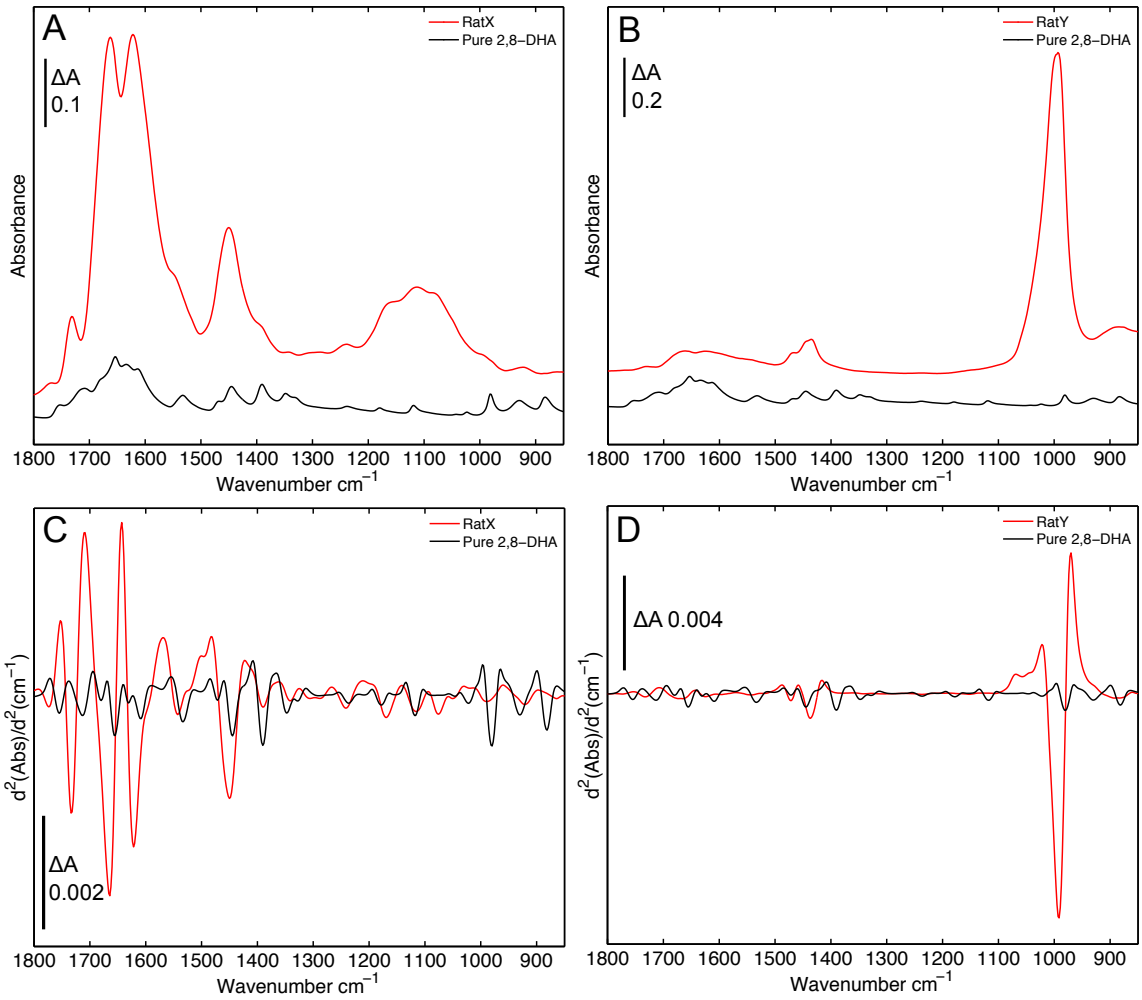


Figure 5.21: IR spectra of urine from rat model of APRTd. *Superimposed IR spectra of dried insoluble fractions of rat model urine samples X (A & C) and Y (B & D) (red) and synthesised 2,8-DHA (black). Top row shows absolute spectra and bottom row shows the second derivative transformations of the absorbance spectra immediately above.*

Although the rat urine spectra differed to the human urine spectra, with RatY having a strong absorbance band at around 1000 cm^{-1} that was absent from human urine spectra, and very little absorbance in other regions of the spectrum (Figure 5.21 B), and RatX having an additional absorbance band at around 1740 cm^{-1}, neither appeared to contain any detectable 2,8-DHA.

5.2.3 Analysis of a transmission spectrum of urine from a suspected APRTd patient

A transmission FTIR spectrum of a urine sample from a suspected APRTd patient was supplied by Michael P Metz, SA Pathology.

Comparison of the transmission IR spectrum of dried, pelleted urine from a patient with suspected APRT-deficiency with the absorbance IR spectrum of 2,8-DHA measured at UCL indicates that this urine does not contain 2,8-DHA. Figure 5.22 shows the comparison of these two samples and reveals the absence of bands that would indicate the presence of 2,8-DHA.

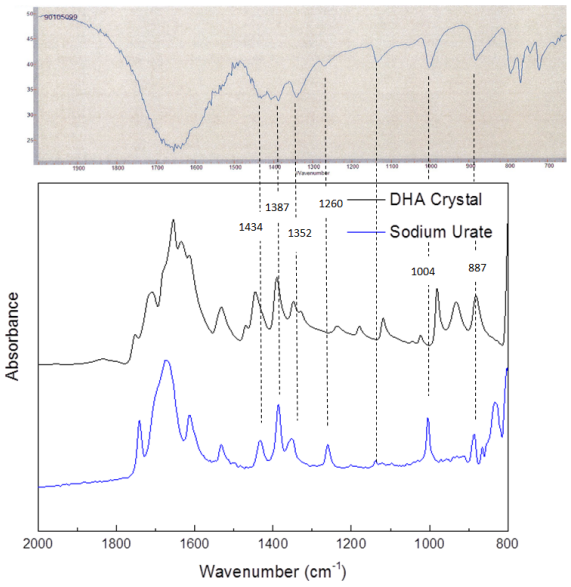


Figure 5.22: Transmission spectrum of urine from a patient with possible APRT-deficiency (top) aligned with spectra of 2,8-DHA and sodium urate (bottom).

The aligned spectra show that the supplied patient urine spectrum has absorbance bands at 1434 cm^{-1}, 1387 cm^{-1}, 1352 cm^{-1}, 1260 cm^{-1}, 1004 cm^{-1} and 887 cm^{-1}, all of which can be seen in the spectrum of sodium urate. The bands in the urine spectrum at 1004 cm^{-1} and 1260 cm^{-1} are clearly absent from the DHA spectrum. In addition, the DHA bands at 1237 cm^{-1}, 1179 cm^{-1}, 1119 cm^{-1}, 980 cm^{-1} and 932 cm^{-1} are absent from the patient urine spectrum. It is concluded that the urine pelleted material has a large urate content but does not contain significant amounts of DHA.

(Note:- the supplied spectrum was taken in transmission mode, whereas ours are shown in absorbance mode; hence the inversion. Also, there is some distortion of bands in the transmission spectra, probably caused by well-known physical phenomena that can affect ATR spectra in certain configurations. It is also likely that the sample was not completely

dried, contributing to the broad absorbance around 1650 cm^{-1}. There may be additional minor component(s) that account for possible additional bands).

5.3 Case study: trimethylaminuria

5.3.1 Introduction to trimethylaminuria

Trimethylaminuria is a hereditary condition in which excess levels of the tertiary aliphatic amine, trimethylamine (TMA), are excreted into bodily fluids, including urine and sweat. This foul-smelling compound results in an unpleasant body odour that has lead to the condition sometimes being referred to as 'fish odor syndrome'. Although not a life-threatening condition, trimethylaminuria can have severe psychosocial consequences and is associated with depression and an increased rate of suicide [148].

TMA is produced by bacteria in the gut from dietary precusors, such as choline, and is usually absorbed rapidly from the gut before being converted to non-odorous TMA N-oxide by the enzyme, flavin-containing monooxygenase 3 (FMO3). TMA N-oxide is then excreted into the urine. Trimethylaminuria results from a defect in this metabolic pathway and in particular, is associated with mutations in the gene that codes for FMO3 [149]. The structures of TMA and TMA N-oxide are shown in Figure 5.23.

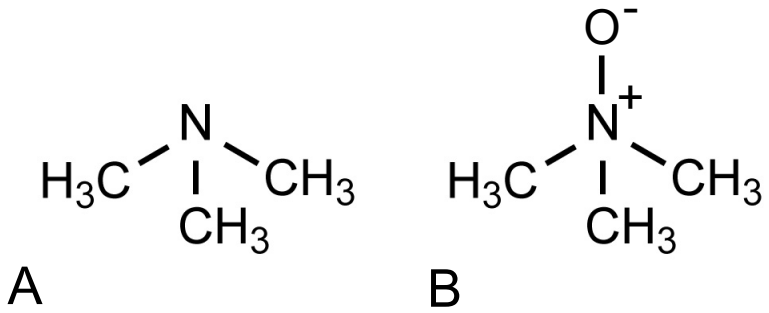


Figure 5.23: *Structures of trimethylamine (TMA) and TMA N-oxide. (A) Structure of TMA. (B) Structure of TMA N-oxide.*

Diagnosis of trimthylaminuria is typically achieved by assessment of clinical symptoms and from the ratio of TMA to TMA N-oxide in urine samples. The identification of the genetic mutations associated with the condition, and the signifiant reduction in the cost of gene sequencing means that genotyping is now also a diagnostic option. However, diagnosis is still problematic due to the variability in the symptoms within and between individuals, and because the disease is associated with several different genetic, and potentially non-genetic causes [148, 149].

The potential of FTIR-spectroscopy as a rapid and sensitive alternative method of detecting TMA and TMA N-oxide in urine samples was investigated. Such a method could have potential uses clinically and experimentally when trying to detect, diagnose and characterise trimethylaminuria.

5.3.2 Vapour and solution phase spectra of trimethylamine

Since TMA is particularly volatile, vapour phase transmission FTIR spectra of TMA solutions were recorded to determine whether clinically relevant concentrations could be detected. Figure 5.24 A shows the vapour spectrum of a 10 mM sample of TMA between 3000 cm^{-1} and 2500 cm^{-1}. The peaks centred around 2823 cm^{-1} and 2774 cm^{-1} are due to the TMA. Figure 5.24 B shows the vapour spectrum of a 1 mM sample of TMA in the same region; the TMA peaks are not detectable at this concentration under the conditions used.

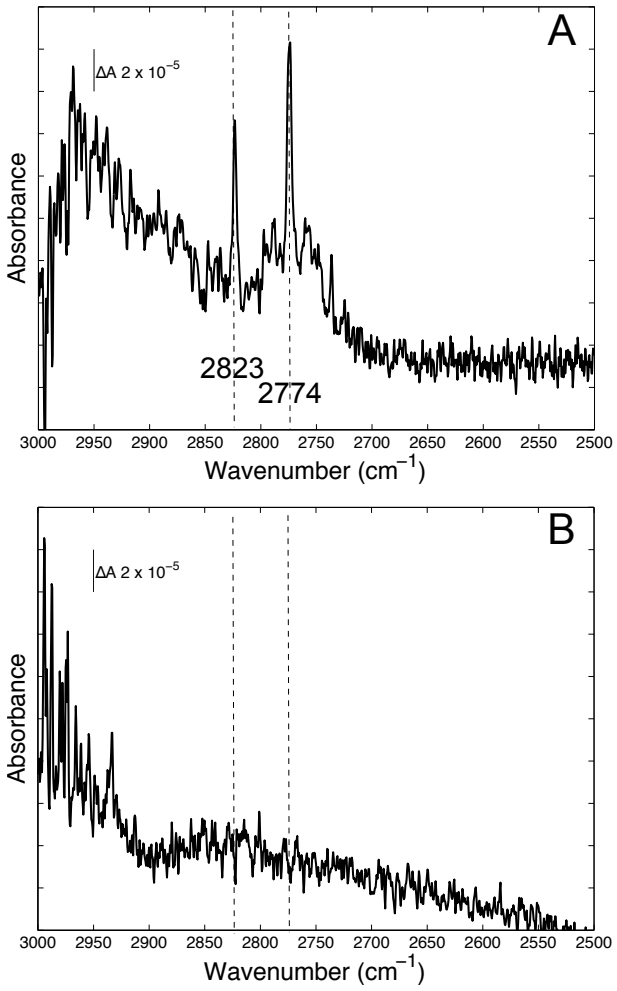


Figure 5.24: Vapour phase transmission spectra of TMA solutions. *(A)10 mM. (B) 1 mM.*

Figure 5.25 A shows the solution spectrum between 1800 and 850 cm^{-1} of 50

mM TMA after subtraction of background buffer and water vapour spectra. Figure 5.25 B shows the same spectrum between 3000 and 2500 cm^{-1}. Whilst the signals are small there is one clear feature at 1487 cm^{-1}. This feature could not be detected in the vapour phase spectra as it falls in a region where water vapour absorbs strongly. The absorbance of water vapour in the vapour phase spectra was at least 100 times greater than the 10 mM TMA absorbance. The features between 1200 and 900 cm^{-1} in the solution spectra are caused by phosphate. Figure 5.25 B reveals that the peaks at 2823 cm^{-1} and 2774 cm^{-1} seen in the vapour phase are not detectable above the intrinsic noise in the solution phase spectrum at this concentration.

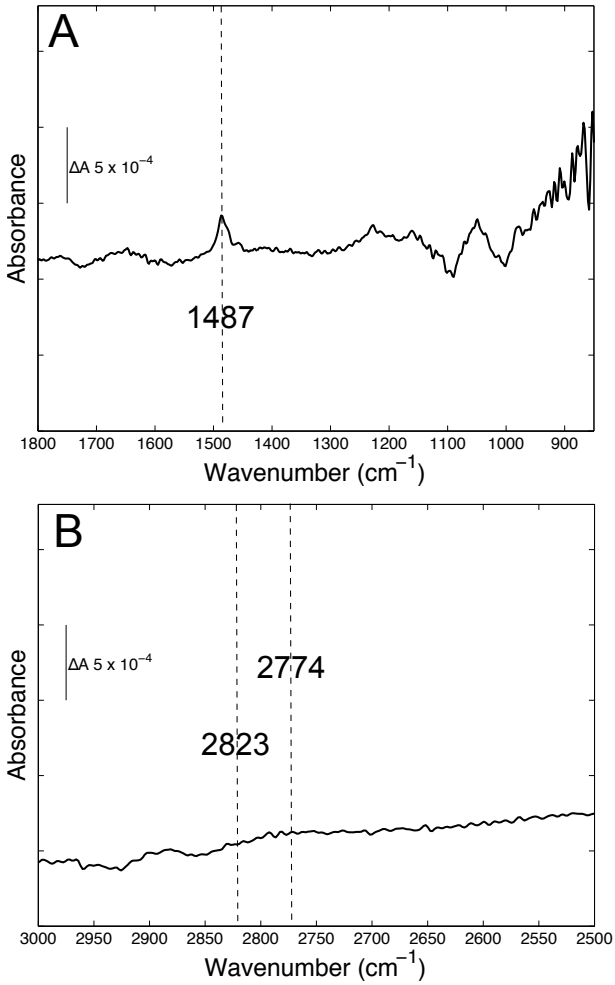


Figure 5.25: Solution ATR spectra of TMA. *50 mM TMA in buffered solution (pH 7) with buffer background and water vapour subtracted. (A) Spectrum between 1800 and 850 cm^{-1}. (B) Spectrum between 3000 and 2500 cm^{-1}.*

Solution phase spectra of non-volatile TMA derivative, TMA N-oxide, were also recorded. Figure 5.26 A & B show the solution spectra between 1800 and 850 cm^{-1} of 25 mM and 5 mM TMA N-oxide respectively, treated as described for TMA. Peaks at

1481, 1466, 1240 and 950 cm^{-1} can be detected at both concentrations. The strong peak at 950 cm^{-1} is likely to be caused by the NO stretch. The negative peak at around 1630 cm^{-1} is likely to be the result of TMA N-oxide interacting with water, altering the water structure and leading to an imperfect subtraction. Again, there are no peaks at 2823 cm^{-1} and 2774 cm^{-1} as can be seen in Figure 5.26 C.

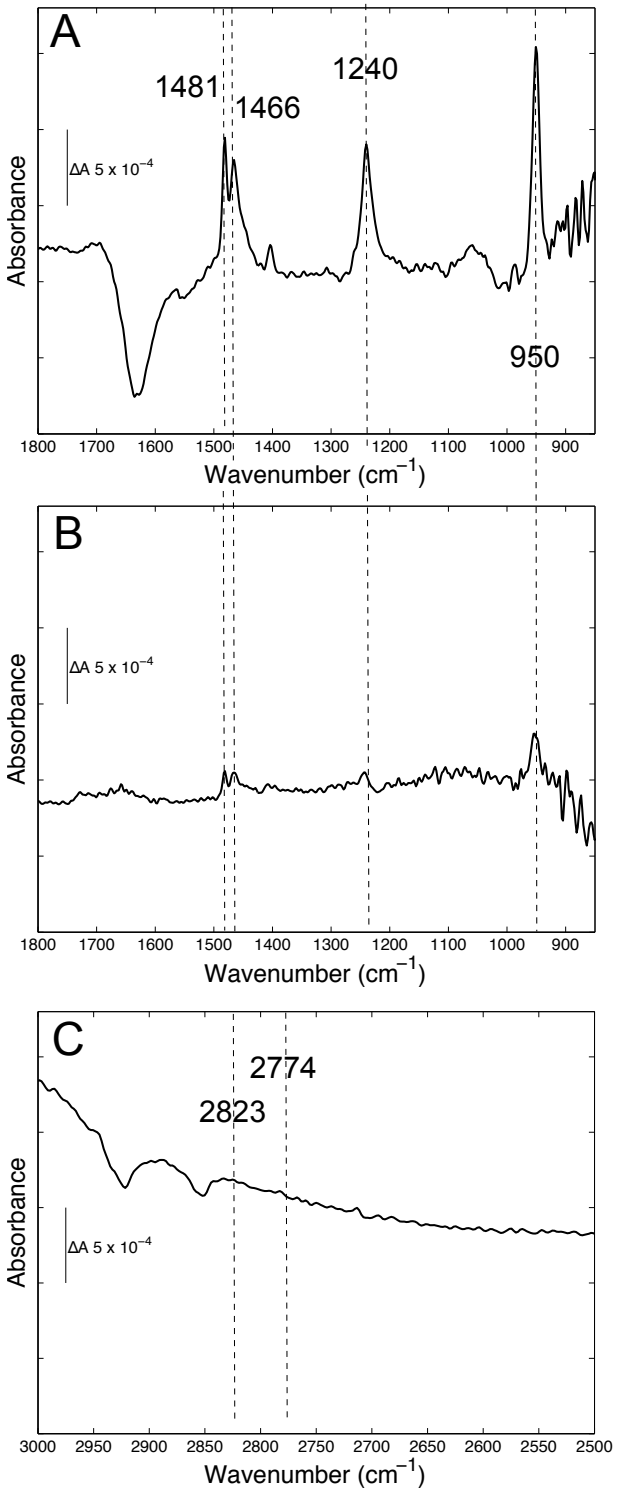


Figure 5.26: Solution ATR spectra of TMA N-oxide. *1800 - 900 cm^{-1} regions of the spectra of 25 mM (A) and 5 mM (B) TMA N-oxide solutions. (C) 3000-2500 cm^{-1} region of 25 mM TMA N-oxide solution. All samples were adjusted to pH7 in phosphate buffer and the same buffer was used as a background during spectra acquisition. A spectrum of pure water vapour was subtracted from each sample spectrum.*

5.4 Discussion and conclusions

The purpose of this study was to assess whether ATR-FTIR spectroscopy could be used as a tool for detecting disease biomarkers in urine. The main focus was to evaluate the technology as a fast, cost-effective method for routine monitoring of urinary cystine. To achieve this goal, we focused on assessing the level of the insoluble cystine, because it can be separated from possible interfering soluble components and because it is this insoluble material that promotes stone formation and causes the cystinuric disease complications.

Since patient fluid intake before urine collection cannot be controlled easily, insoluble cystine load was also compared to urinary creatinine concentrations to take account of urine dilution. Both urea and creatinine can be quantitated using IR spectra of whole (i.e. undried) urine samples [49, 91]. Creatinine was determined with a simple algorithm that deconvoluted the overlapping bands of urea and creatinine that dominate the 1510-1445 cm^{-1} region of whole urine spectra (Figures 5.1 - 5.4). These ATR-FTIR derived values were validated by comparison with values obtained by the clinical standard Jaffe reaction. A reasonable correlation between values determined by the two methods was found, although the Jaffe reaction tended to give slightly higher values. This could be a result of additional chromogens in urine that interfere with the Jaffe reaction method [143, 150, 151]. However, there were also two significant outliers showing large discrepancies between the creatinine measured by the two methods for reasons that have not been resolved. The remaining data points fell within the range defined by the 95% confidence interval (± 1.96 standard deviations of the mean bias) [99]. For 67% of the samples the difference between the two methods was less than 2 mM creatinine and for 81% of samples the difference was less than 3 mM creatinine.

Sufficiently accurate quantitation of the insoluble cystine in these dried samples was possible by comparison of the height of its 1296 cm^{-1} band in second derivative to a calibration curve of the same band in spectra of dried, pure cystine suspensions. This band was chosen because there is no strong absorber at this frequency in whole urine or in other insoluble components that might be present in urine from patients with other kidney conditions (including oxalate, urate and dihydroxyadenine). A line of best fit through the origin (Figure 5.9 B) was linear to at least 3 mM cystine and all but one (where the cystine concentration was 3.1 mM) of the cystinuric samples measured fell within this range.

A good correlation was found between the cystine levels measured by the FTIR and clinical ion exchange chromatography methods. Notably, the FTIR method measures

only the insoluble cystine in a sample, whilst the clinical method aims to measure total cystine without specifically reporting on insoluble cystine. In order to compare the two methods, a 1 mM offset is applied to the FTIR values to account for the additional soluble component included in the clinical values. However, the ability to assess only the level of insoluble cystine in urine may be most useful clinically as it is this insoluble fraction that leads to disease complications.

Since the cystinuric patients included in this study were advised on fluid intake and diet, and some were receiving drug treatment (Table 5.3), it could be expected that some of the urine samples contained no insoluble cystine. Indeed, four cystinurics with no detectable urinary insoluble cystine (P5, P13, P20 and P23) and three patients (P7, P8, P11) with low insoluble cystine were receiving cystine-decreasing treatments. An additional patient (P14) urine sample contained low insoluble cystine despite no drug treatment, but the creatinine concentration was also very low indicating highly dilute urine. P3 also had no detectable insoluble cystine, but was not on drug treatment and the creatinine concentration was not particularly low, suggesting that this patient suffered from only mild cystinuria. In contrast, high levels of cystine were recorded in P1, P6, P15, P18 and P19; these patients were either not receiving drug treatment for cystinuria (P6, P15 and P19) and/or had concentrated urine as judged by relatively high creatinine concentrations (P1, P15 and P18). When assessing stone formation risk, a more useful clinical parameter than absolute cystine concentration may therefore be the cystine/creatinine ratio (Figure 5.27). For example, P1 and P18 had comparatively high insoluble cystine concentrations, but P1 had a much higher cystine to creatinine ratio, suggesting that this patient might have a more severe case of cystinuria.

Table 5.3: Data on cystinuric patients used in study. *Cystine and creatinine data are those determined by ATR-FTIR methods described here (od = once daily, bd = twice daily, tds = thrice daily, qds = four times per day, nocte = at night, mane = in morning).*

	Age (years)	Sex	Current cystinuria treatment	Mean [Creatinine] (mM)	Mean [Cystine] (mM)	Cystine : creatinine (μM/mM)
P1	19	Female	Potassium citrate (1 tablet bd)	9.9	2.82	286
P2	22	Male	None	7.6	0.71	94
P3	33	Male	None	8.4	0.00	0
P5	36	Female	Tiopronin (250 mg tds)	1.1	0.00	0
P6	22	Male	None	7.3	1.66	228
P7	21	Male	Tiopronin (250 mg tds) Potassium citrate (2 tablets qds) Effercitrate (2 tablets qds)	2.5	0.12	49
P8	54	Female	Penicillamine (750 g nocte) Penicillamine (250 mg mane)	3.8	0.12	32
P10	29	Male	Potassium citrate (1 tablet bd) Tiopronin (500 mg nocte and 250 mg mane)	8.1	1.62	200
P11	59	Female	Tiopronin (250 mg bd)	2.5	0.02	8
P12	64	Male	Potassium citrate (10 ml tds)	8.9	1.14	128
P13	32	Female	Potassium citrate (1 tablet bd)	1.0	0.00	0
P14	59	Male	None	0.0	0.11	undefined
P15	35	Female	None	12.8	3.13	245
P16	36	Female	Tiopronin (250 mg tds)	5.6	0.51	91
P17	22	Male	Potassium citrate (1 tablet bd) Tiopronin (200 mg bd)	7.0	0.84	121
P18	67	Female	Tiopronin (500 mg bd) Potassium citrate (2 tablets bd)	20.6	1.69	82
P19	34	Female	None	8.6	1.51	176
P20	48	Female	Sodium bicarbonate (1 g nocte) Penicillamine (500 mg od)	3.4	0.00	0
P21	52	Male	Potassium citrate (10 ml tds)	9.3	1.28	138
P22	54	Male	None	8.2	0.14	17
P23	63	Female	Penicillamine (250 mg tds)	6.2	0.00	0
P24	34	Female	None	1.6	0.22	140

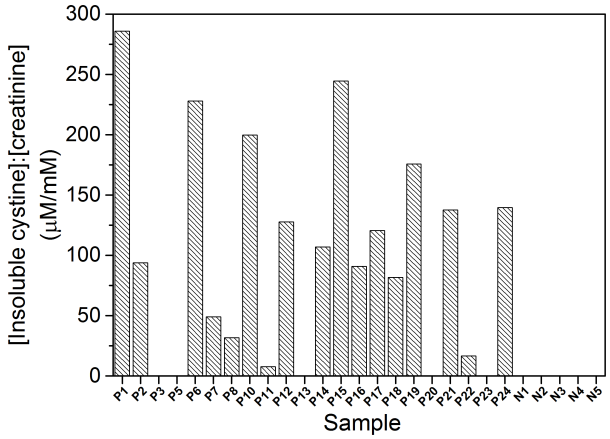


Figure 5.27: Urinary cystine to creatinine ratios. *Bar graph showing the insoluble cystine:creatinine (μM:mM), both determined by the ATR-FTIR spectroscopy methods, for the 22 cystinuric samples and 5 controls.*

In summary, we show here that cystine can be quantitated from FTIR spectra of the dried insoluble fractions of urine samples without any requirement for chemical manipulation. A single band at 1296 cm^{-1} was selected for a simple quantitation; however, further enhancement of sensitivity and specificity for cystine could be achieved if required by inclusion of additional cystine bands (e.g. those at 845 and 775 cm^{-1}) in the analytical algorithm or by other types of quantitation methods. The protocol could also be adapted for use with a high throughput, automated sampling system, for example on silicon plates with transmission mode FTIR spectrometer, thus providing a rapid means for diagnosing cystinuria that could be labour- and cost-effective. Such an ability to accurately measure urinary cystine concentrations on a regular basis in patients with cystinuria would be of significant benefit both for diagnostic screening and for monitoring the effectiveness of treatments.

A smaller pilot study was performed to determine whether the same approach could be used to detect DHA in urine as a diagnostic assessment for APRT-deficiency. The region between 850 cm^{-1} and 1000 cm^{-1} is likely to be particularly useful for detection and quantitation of insoluble DHA in human urine samples, because the bands are relatively strong and fall in a region of the urine spectrum that is less dense with bands from other components.

The analysis of the 2,8-DHA stone sample and urine sample A from an APRT-deficiency patient indicates that FTIR-spectroscopy can be used to detect multiple characteristic bands of DHA in urine samples. Pelleting and drying the samples before measurement, coupled with analysis of second derivative spectra increases the sensitivity of the method. However, it was not possible to confidently detect 2,8-DHA in the other

5 urine samples B-F. This may be because those samples did not contain any 2,8-DHA or contained only very low quantities beyond the detection sensitivity of the method described here. It is also possible that patient C is not APRT deficient.

Comparison of urine samples B-F with other known stone-forming compounds did not provide any further insight into the composition of the insoluble material in the urine. Some of the absorbance bands arise from residual urea and creatinine that cannot be easily separated from the insoluble pellet after centrifugation. Urea has strong absorption bands at 1468 $cm^{-1}$and between 1715 - 1580 cm^{-1} in absolute spectra. Creatinine has a strong absorption band at 1492 cm^{-1} in absolute spectra. Spectra of all of the human urine samples, both insoluble and soluble fractions, showed bands associated with urea and creatinine. Urine samples E and F are also quite likely to contain significant protein in the insoluble fractions (Figure 5.18 F&G) based on the presence of the amide II absorption band at around 1550 cm^{-1}.

Development into a quantitative method is feasible with generation of appropriate calibration curves and such analyses with a simplified bench top instrument within the clinic should be possible. However, more urine samples would have to be analysed to determine whether the sensitivity of the method is adequate for clinical application.

Comparison of the synthesised 2,8-DHA FTIR spectrum with a query spectrum of urine from a patient with suspected APRT-deficiency (Figure 5.22) reveals that this is unlikely to be the correct diagnosis since no significant 2,8-DHA was detected. This spectrum instead indicated significant levels of urate and this could itself be a cause of stone formation [152].

Analysis of the insoluble fractions of urine from a rat model of APRTd also did not show any evidence of the presence of 2,8-DHA. There were, however, several additional bands that are not seen typically seen in healthy human urine. Further investigation would be needed to identify the origin of these bands and to develop a more comprehensive analytical protocol for rat urine samples.

As well as kidney stone constituents, other compounds can be found in human urine that are diagnostic for other types of disease. FTIR spectroscopy can be a useful tool for detecting those compounds that are difficult to measure by other means. However, many of these are soluble and so cannot be measured using the procedures described for cystine and DHA. Here, the possible use of FTIR spectroscopy was also applied to the detection of soluble trimethylamine (TMA) in solutions to determine whether clinically relevant levels of TMA or TMA N-oxide could be detected in either vapour or solution phases.

It was possible to detect TMA in the vapour phase at 10 mM and in solution at 50 mM using FTIR spectroscopy. Lower concentrations can be detected under vapour phase conditions because long path lengths can be used, and because absorbance bands are much sharper in the vapour phase compared to the solution phase. Under the experimental conditions used in this preliminary study, the expected limit of detection of TMA in the vapour phase would be around 5 mM. However, it may be possible to detect lower concentrations with optimised conditions, for example by using mirrors to increase the pathlength of the IR light through the vapour or by adjusting the instrument settings (e.g. increasing the number of background/sample scans and adjusting the scanner velocity). It is unlikely that TMA can be detected in the solution phase at physiological concentrations because the signals are weak and the peaks are broadened through interaction with the aqueous environment. Furthermore, the symmetrical nature of TMA means there are few infrared active vibrational modes.

TMA N-oxide is non-volatile so cannot be studied by vapour phase FTIR spectroscopy. However, the N-oxide group is strongly infrared active could be detected using solution phase ATR-FTIR spectroscopy. TMA N-oxide produced a distinctive IR spectrum and could be detected down to 5 mM, although detection of lower concentrations in solution would be challenging. Drying the samples would increase the signals in pure solutions, and may offer a way of increasing the signals in urine samples.

A summary of the diagnostic methods for cystinuria, APRTd and trimethylaminuria, including the diagnostic materials and potential detection ranges, is presented in Table 5.4.

Table 5.4: Summary of diagnostic methods developed using FTIR spectroscopy for detection of urinary disease biomarkers.

Disease	Diagnostic material	Method used	Detection range	Physiological range
Cystinuria	Cystine	ATR-FTIR spectroscopy of dried insoluble material. Area 1296 cm^{-1} band in second derivative spectra calculated.	0.1-3+ mM insoluble cystine	1-4+ mM total cystine.
APRTd	2,8-DHA	ATR-FTIR spectroscopy of dried insoluble material. Identification of absorbance bands at 1390 , 1117, 979, 932 and 881 cm^{-1} in second derivative spectra.	Not established. Detection limit likely to be at least 0.1 mM.	Solubility limit is 0.009 mM DHA.
Trimethylaminuria	TMA/TMA N-oxide	Identification of absorbance bands at 2823 and 2774 cm^{-1} in vapour phase transmission spectroscopy of TMA. Identification of absorbance bands at 1481, 1466, 1240 and 950 cm^{-1} in solution phase ATR-FTIR spectroscopy of TMA N-oxide.	4-5+ mM TMA in vapour phase. 5+ mM TMA N-oxide in solution phase, likely to be lower if dried.	0.08-0.8 mM TMA.

Chapter 6

ATR-FTIR spectroscopy as a point-of-care diagnostic tool for detecting disease biomarkers in fresh urine samples

Author contributions: this chapter has been published in Proceedings of SPIE 9332, Optical Diagnostics and Sensing XV: Toward Point-of-Care Diagnostics. All data analysis was performed by Katherine Oliver. All figures in the publication were prepared by Katherine Oliver. Urine spectral data was collected by Faith Matjiu, Cameron Davey and Katherine Oliver. The paper was written by Katherine Oliver with contributions from Peter R. Rich. Shabbir H. Moochhala and Robert J Unwin offered clinical guidance. Additional Figures have been included.

6.1 Introduction

Urine samples contain a wide range of constituents that can provide valuable diagnostic markers of disease states [108]. A range of clinical tests for many of these are used routinely, although there is still a need for additional simple, rapid and high throughput methods.

As discussed in the previous chapter, ATR-FTIR spectroscopy provides a simple and fast means of recording spectra of various samples that are rich in information that might be useful for diagnosing some diseases [3, 49, 79–81, 153, 154]. Until recently, the spectrometers required for such analyses were complex and expensive. Since the advent of commercially available, low maintenance, benchtop instruments, coupled with advanced and automated methods of data analysis, new possibilities for clinical applications have emerged. These newer machines are designed for continuous use for basic mid-IR analysis and allow quick, easy and reliable FTIR analysis with extremely small footprints. Unlike research-grade spectrometers, benchtop spectrometers operate with room temperature detectors and permanently aligned interferometers that guarantee precise data acquisition without adjustment.

Biofluids are particularly attractive sample types for bedside disease diagnosis since they can generally be obtained quickly and non-intrusively. Mitchell *et al.* (2014) [155] reviewed a number of studies in which vibrational spectroscopy was used to identify biomarkers in various biofluids (e.g. serum, whole blood and cerebrospinal fluid) for a range of diseases including cancers, diabetes and neurodegeneration. They hypothesised that a microfluidic device in combination with ATR-FTIR spectroscopy could be designed to provide bedside analysis of biofluids.

ATR-FTIR spectroscopy has also been used to analyse urine composition. For example. Heise *et al.* (2001) [81] and Hoşafçi *et al.* (2007) [49] used partial least squares (PLS) analysis to quantitate urea, creatinine, uric acid, phosphate and sulphate from

urine sample spectra.Such methods could feasibly be translated to a bedside ATR-FTIR device for point-of-care diagnostics.

Many potential disease biomarkers have been identified in urine using techniques other than vibrational spectroscopy, including, biomarkers for early chronic kidney disease (CKD) [156], acute kidney injury (AKI) [157, 158], bladder cancer [159, 160] and even lung cancer [161]. Many of these biomarkers are proteins that would be difficult to distinguish in FTIR spectra. However, it might be possible to detect some of these biomarkers from ATR-FTIR spectra of urine samples recorded on benchtop machines, for example, lipoprotein and glycoprotein for CKD [156], total protein for AKI [157] and hyaluronic acid for bladder cancer [159, 160]. This would enable a diverse range of diseases to be diagnosed or screened rapidly and non-invasively using a single diagnostic test.

The purpose of this study was to explore further the use of ATR-FTIR spectroscopy as a point-of-care diagnostic tool. A benchtop spectrometer was installed in the Kidney and Urology Centre at the Royal Free Hospital (RFH), London (UK) to explore the feasibility of operating such an instrument in a clinical setting and assess whether sufficiently high quality spectral data can be obtained. Two nurses in the clinic were asked to record spectra of wet urine samples collected from as many patients seen in the clinic as possible, so as to generate a large database of urine spectra from a variety of patients without targeting a particular kidney-related disease or patient demographic.

6.2 Development of protocols for measuring urine samples on the benchtop spectrometer

Since the benchtop spectrometer was to be installed in a hospital clinic and operated by non-specialists with limited training (mainly staff nurses) it was important to ensure that good quality reproducible data could be generated with relative ease. A crucial aspect was that the staff recording the measurements could do so with minimal interruption to their normal working day. Therefore, the protocols for measuring the urine sample spectra were developed to be as simple and quick as possible.

Problems associated with drying urine samples on to the ATR prism had been previously observed and reported in earlier chapters. These include: hydration effects; partitioning of urine components; variations in drying times; saturation of the IR signal at high urine concentrations; and the need for very precise sample volumes. Furthermore, a protocol involving drying the urine samples would require additional equipment,

training and time. Therefore, it was decided that urine samples would be recorded wet and unprocessed.

A short manual was produced for staff operating the benchtop spectrometer describing the protocol for recording the urine sample spectra (Appendix A). A macro was also written in OPUS 7 which enabled staff to record background and sample measurements without viewing or adjusting any software/hardware settings. Buttons were placed on the main OPUS interface that could be clicked to call either the background measurement or sample measurement macros (Figure 6.1). Easy-to-follow instructions were then displayed on screen.

Staff were asked to start by thoroughly cleaning the prism with distilled water followed by absolute ethanol using a lint-free tissue. Once the ethanol had completely evaporated (3-5 seconds) a background measurement was recorded (before each sample measurement, or every 30 minutes if recording several samples). Once the 'RECORD SPECTRUM' option was selected a text box appeared with the option to input a sample name and any additional details. Here, staff were asked to enter information about the main diagnosis associated with the sample. A live spectrum of the blank prism was displayed and this was examined to ensure the prism was clean. Around 100 μL of fresh unprocessed liquid urine was then pipetted directly onto the prism so that it covered the entire prism and the measurement was started. The urine samples were removed from the prism as soon as the measurements had finished. The macro automatically unloaded the spectrum from the OPUS interface and saved it to a specified directory on the computer.

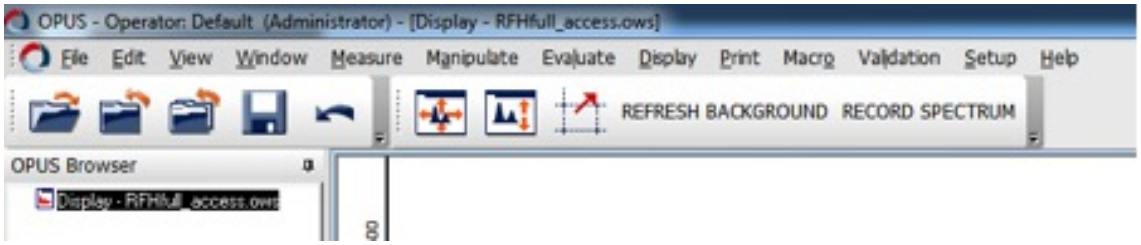


Figure 6.1: Simplified buttons for evoking macros in OPUS 7.

6.3 Final dataset after quality control

All samples were collected from patients of the Kidney and Urology Centre at the RFH. Age, sex, primary diagnosis (associated with visit to the clinic), confounding diagnoses and current medications, as well as estimated glomerular filtration rate (eGFR) were recorded for each patient. Two hundred and ninety-five spectra of wet urine sam-

ples from 256 patients were recorded. To determine whether a sample should be taken forward for further analysis the signal size of the urea absorbance band at 1468 cm^{-1} was considered along with the signal to noise ratio (SNR) of the spectrum. To be included in the analyses the intensity of the absorbance at 1468 cm^{-1} after water subtraction and before normalisation had to be above 1.5 $x10^{-3}$ absorbance units. The SNR cut-off for inclusion in the analyses was 1.3 at 1468 cm^{-1}. After quality control, 206 spectra from 190 patients were taken forward for subsequent analyses. The age and sex distribution of these 190 patients is summarised in Table 6.1.

Common diagnoses included:

- Stone disease (calcium phosphate/calcium oxalate, urate and cystine stones) as well as hyperoxaluria, hypercitraturia and hypercalciuria, which are all associated with increased risk of stone disease;
- Various diseases associated with kidney damage including polycystic kidney disease and renal cysts, chronic and familial kidney disease (CKD), end stage renal failure (ESRF), nephritis and proteinuria;
- Recurrent urinary tract infections (UTI);
- Renal transplants.

An additional 42 urine samples were collected from healthy volunteers (no known medical conditions), including staff at the RFH and individuals from University College London (UCL) and Imperial College London. In most cases these urine samples were donated anonymously and so age and sex information is not available.

Table 6.1: Ages and sex of patients donating urine samples to study. *An additional 42 urine samples from healthy volunteers were collected but age and sex information is not available for most of these individuals.*

	Age				**Sex**	
	18-30 yrs	**31 - 50 years**	**51 - 70 years**	**≥71 years**	**Male**	**Female**
Urine samples collected at RFH	22	83	76	9	116	74

6.4 Identification of spectral variation and comparison to model compound spectra

Examination of the individual pre-processed urine spectra showed that whilst the overall spectral band pattern was similar across all samples, there were a number of prominent differences as well as subtle changes in the shape or position of some bands. Some of these changes fall in regions that are associated with specific biochemical compounds such as proteins or sugars and can therefore be correlated with known disease phenotypes. For example, the presence of an amide II band at around 1550 cm^{-1} indicates the presence of protein in the urine and suggests the patient may have kidney disease. The fingerprint regions of seven example spectra showing several variations are plotted in Figure 6.2.

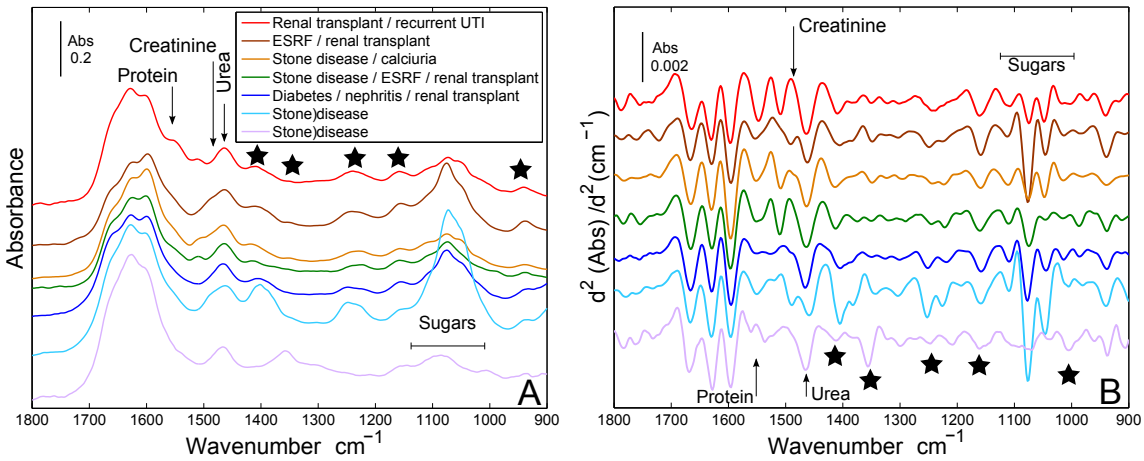


Figure 6.2: Examples of urine spectra recorded in the RFH clinic. *Absorbance (A) and second derivative (B) spectra showing specific biochemical components including urea, creatinine, protein and sugars. Stars represent spectral features that vary between individual spectra and may be of diagnostic interest.*

To confirm that the protein amide II, urea, creatinine and sugar absorbance bands identified in Figure 6.2 are correctly assigned, the urine spectra were compared to some model compound spectra. In Figure 6.3, three example urine spectra are plotted along with spectra of pure urea, creatinine, protein (BSA, as a model protein), glucose, lactose, potassium phosphate (pH 7) and potassium sulphate (pH 7).

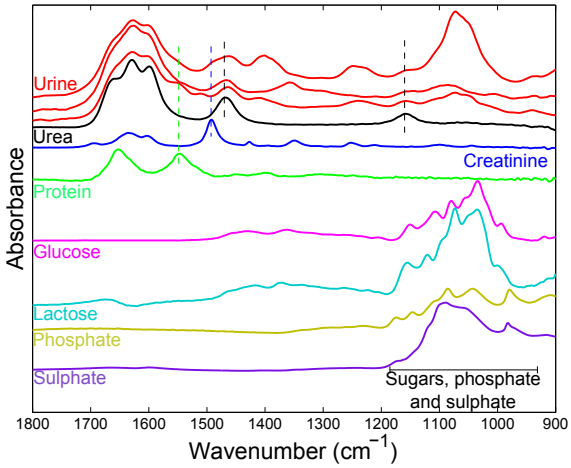


Figure 6.3: Comparison of three example urine spectra with model urinary compound spectra.

6.5 Comparison of all renal-related disease urine spectra with healthy urine spectra

As an initial comparison, the urine spectra collected from diseased patients at the RFH were compared to set of urine spectra from healthy individuals to determine whether there were any common spectral features associated with renal-related diseases. Figure 6.4 A shows an average of the 206 absorbance spectra of diseased urine samples compared with an average of 42 absorbance spectra of healthy urine samples. The second derivatives of these absorbance spectra were also calculated to give better resolution of shoulders and remove broad baseline artefacts. Figure 6.4 B shows the averaged second derivative spectra of all RFH disease samples and of the 42 healthy (control) urine samples, while Figure 6.4 C shows the individual second derivative spectra. Peaks of potential diagnostic interest have been highlighted with the arrows and numbered 1 to 17; where two arrows are indicated for a single peak, both peak intensity and peak shift were analysed.

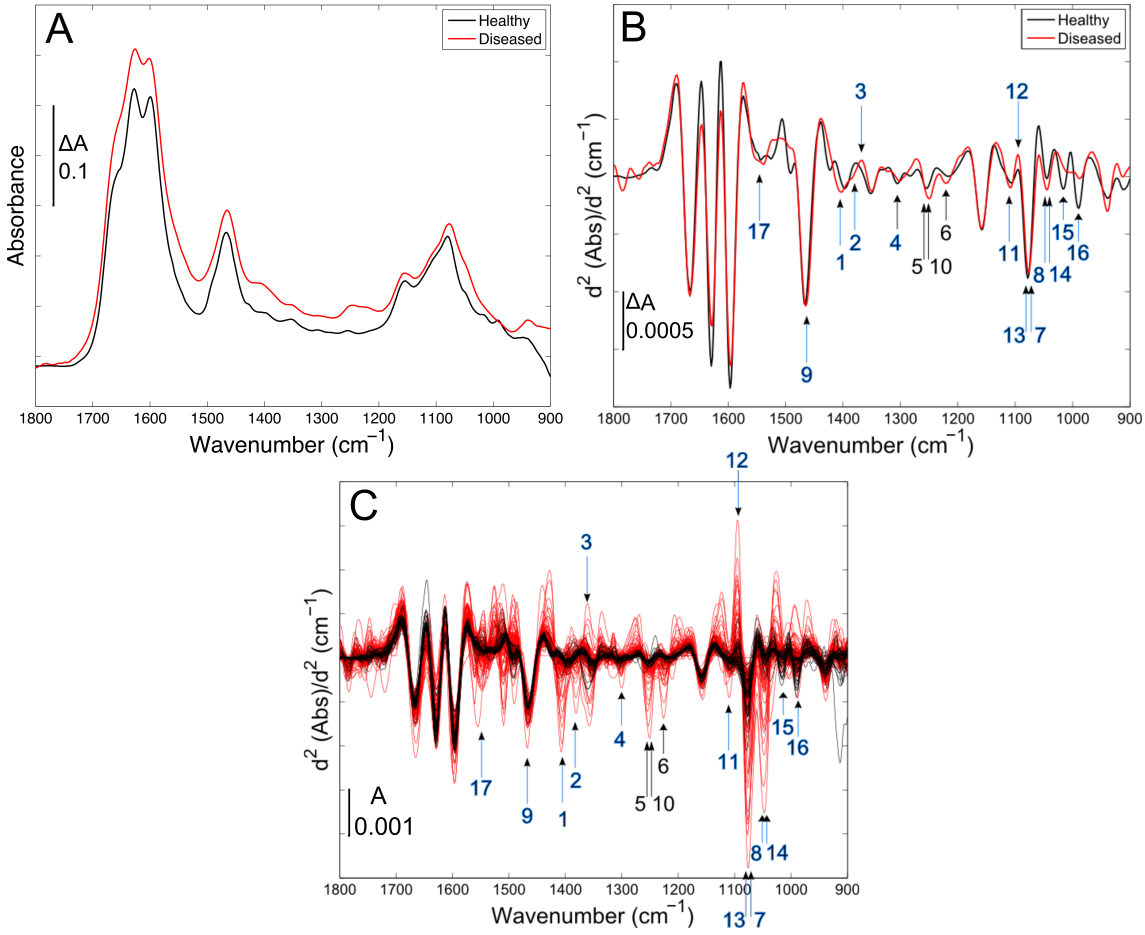


Figure 6.4: Healthy vs. diseased urine spectra recorded in RFH clinic. *All RFH disease urine spectra (red) versus healthy urine spectra (black). (A) Average of all RFH disease urine absorbance spectra and of all healthy absorbance spectra. (B) Second derivative transformations of averages shows in A. (C) Individual second derivative spectra. Blue arrows represent features found to be significantly different between the disease and healthy groups at significance levels of at least 0.05. Black arrows represent additional features that showed variation but were not statistically significantly different between the two groups. All spectra were normalised to the area of their urea/creatinine band range at 1468 cm^{-1} between 1487 and 1436 cm^{-1}.*

Some specific components can be identified from these spectra. Absorbance bands of urea at ~1668, ~ 1630, ~1595, ~1468 (feature 9) and 1157 cm^{-1} can be clearly seen, along with creatinine bands at ~1492, ~1426 and 1252 cm^{-1}, which can be seen in several of the urine spectra. Absorbance bands of phosphate can also be correlated with the urine spectra at ~1157, ~1076 (features 7 and 13), ~991 (feature 16) and 939 cm^{-1}. The other absorbance bands and potential diagnostic features have not been assigned but can be be correlated with general compounds such as sugars, proteins and lipids.

A peak area analysis of each of the features indicated in Figure 6.4 B & C was performed. A simple comparison analysis was performed to determine whether there

was a significant difference between the mean areas of the healthy and RFH disease urine spectra for each of the 17 peaks. Since the data were not normally distributed, according to Shapiro-Wilk normality tests performed in Origin 8.6, the Mann-Whitney U test was applied. The peaks that were statistically significantly different between the healthy and RFH disease samples, at a significance level of 0.05, are indicated with blue arrows on Figure 6.4 B. Only four of the 17 features analysed did not show a significant difference between the two groups; these are indicated by black arrows on Figure 6.4 B. Table 6.2 details the p-values for each feature comparison and indicates which peaks are significantly different between the healthy and RFH diseased samples at significance values of 0.05 and 0.01.

Table 6.2: p values for comparison of 17 spectral features between healthy and diseased samples. ** indicates features for which there was a difference between the two groups at a significance level of 0.05 and ** indicates differences at significance level of 0.01.*

Feature number	p-value
1	0.99**
2	0.99**
3	0.99**
4	0.96*
5	0.49
6	0.82
7	0.94
8	0.99**
9	0.99**
10	0.65
11	0.99**
12	0.97*
13	0.99**
14	0.99**
15	0.99**
16	0.99**
17	0.99**

Since the standard deviation for any single peak for this dataset was large and there was considerable overlap in the peak area distributions between the healthy and

RFH disease samples, it was useful to combine several peaks in an attempt to improve separation of the data. Scatter plots of the areas of two or three features were used to see if there was any clustering of samples based on their diagnoses. Figure 6.5 shows a selection of 2D and 3D scatter plots of features that gave the best separation.

The three different spectral feature combinations, shown in Figure 6.5, produced clear clustering of the healthy samples away from the RFH disease samples, with a small region of overlap. This suggests that it might be possible to identify patients with kidney-related diseases based on infrared spectra of their urine.

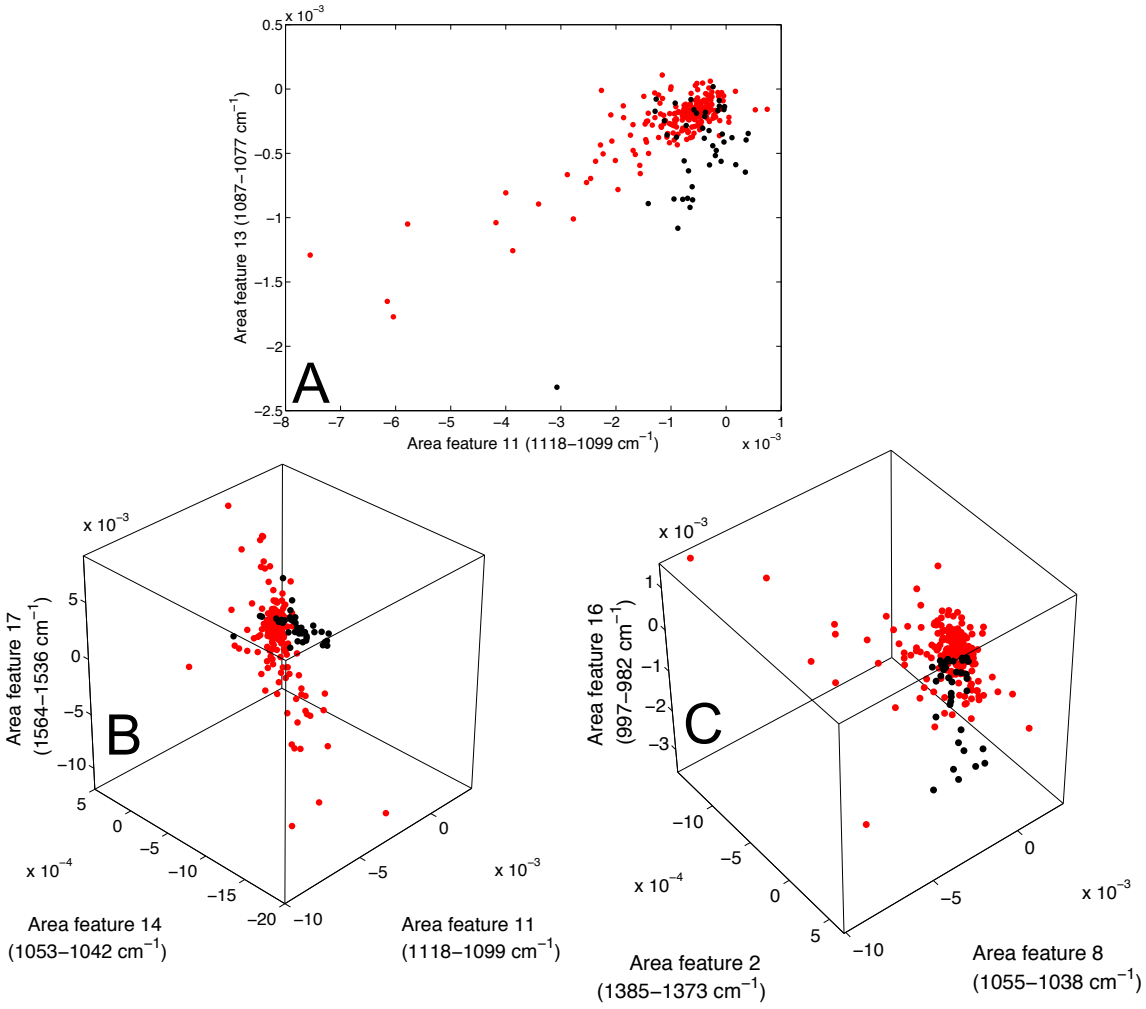


Figure 6.5: Healthy vs. diseased urine scatter plots. *(A) Two-dimensional and (B & C) three-dimensional scatter plots comparing the areas of spectral features found to be significantly different between the disease (red) and healthy (black) urine spectra. Each data point represents an individual spectrum and has been plotted based on the calculated areas of features specified.*

6.6 Spectral identification of urine from patients with abnormal glomerular filtration rate (GFR)

6.6.1 Patients with estimated GFR below 60 mL/min (CKD stage 3+)

The RFH urine sample spectra shown here represent patients with a variety of diseases. Glomerular filtration rate (GFR) can be used as a measure of kidney function. Abnormal GFR can indicate kidney damage arising from a variety of underlying diseases. Since GFR is difficult to measure directly it is often estimated from serum creatinine levels. To determine whether any spectral change can be correlated with specific clinical indicators of kidney damage the RFH urine spectra were grouped according to estimated GFR (eGFR). The threshold corresponding to normal eGFR is 60 mL/min, although most healthy normal individuals would be expected to have a GFR > 90 mL/min. Patients with eGFR < 60 mL/min are classed as having stage 3+ chronic kidney disease (CKD3+) [162, 163]. The urine spectra of these patients were compared with spectra of healthy urine samples, and to a second disease group of urine spectra from patients with eGFR >60 mL/min.

The averaged second derivative spectra of these three groups (healthy, CKD3+ and eGFR>60) are plotted in Figure 6.6. The features that were previously found to vary between healthy and diseased urine spectra were also found to be the most relevant for comparing these groups. The areas of the 17 features for each of the spectra in the CKD3+ group were compared with the areas of the corresponding peaks of the eGFR>60 (normal) group and the healthy group using Mann Whitney U analysis. The features that showed significant differences at a significance level of 0.05 are indicated by arrows in Figure 6.6. Table 6.3 details the p-values for each feature comparison and indicates which peaks are significantly different between the healthy and diseased samples at significance levels of 0.05 and 0.01.

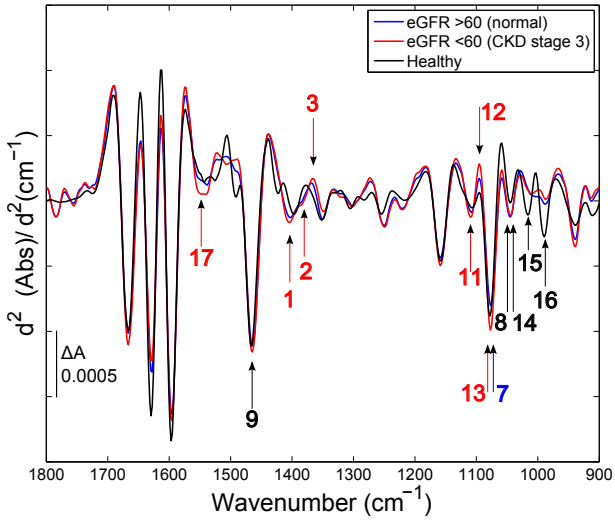


Figure 6.6: CKD3+ patient urine spectra. *Averaged second derivative urine spectra of all healthy (black), all CKD3+ (red, abnormal) and all eGFR>60 (blue, normal) groups. Red arrows represent features found to be significantly different between both the healthy and CKD3+ groups and the eGFR>60 and CKD3+ groups at significance levels of at least 0.05. Black arrows represent additional features that showed significant differences between CKD3+ and healthy groups at significance levels of at least 0.05, but not between CKD3+ and eGFR>60 groups. The blue arrow represents a feature that showed a significant difference between CKD3+ and eGFR>60 groups at a significance level of at least 0.05, but not between CKD3+ and healthy groups. All spectra were normalised to the area of their urea/creatinine band range at 1468 cm^{-1} between 1487 and 1436 cm^{-1}.*

Table 6.3: p values for comparison of 17 spectral features between healthy and CKD3+, and between CKD3+ and eGFR>60 samples. ** indicates features for which there was a difference between the two groups at a significance level of 0.05 and ** indicates differences at significance level of 0.01.*

Feature number	p-value CKD3+ vs. healthy	p-value CKD3+ vs. eGFR>60
1	0.99**	0.95*
2	0.99**	0.97*
3	0.99**	0.99**
4	0.61	0.89
5	0.79	0.36
6	0.59	0.03
7	0.32	0.99**
8	0.25	0.76
9	0.99**	0.04
10	0.07	0.41
11	0.99**	0.99**
12	0.99**	0.99**
13	0.99**	0.99**
14	0.99**	0.24
15	0.99**	0.06
16	0.99**	0.67
17	0.99**	0.99**

Good separation of the healthy and CKD3+ groups could be achieved by plotting the areas of three selected features in a 3D scatter plot; comparison of features 1, 15 and 17 gave the best separation of the two groups (Figure 6.7). Feature 17 corresponds to the amide II vibration of protein and the presence of this band in a spectrum indicates that the urine contained protein. Since proteinuria is an indicator of kidney damage, this observation validates the use of feature 17 in discriminating between healthy urine and urine from patients with abnormal GFR.

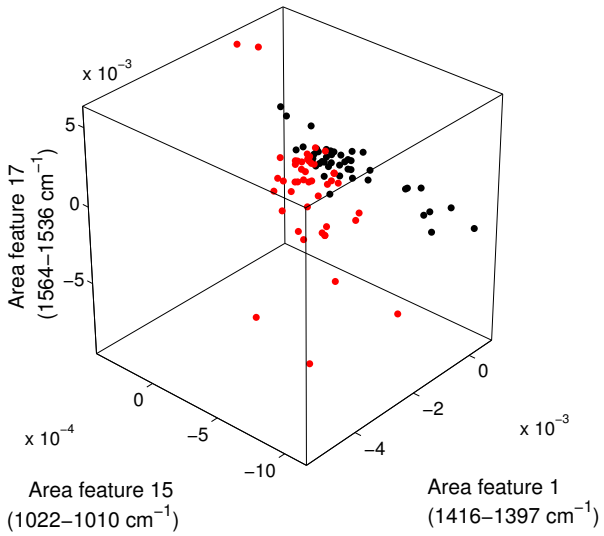


Figure 6.7: CKD3+ vs. healthy scatter plot. *Three-dimensional scatter plot comparing the areas of 3 spectral features found to be significantly different between the CKD3+ (red) and healthy (black) urine groups.*

For comparing the eGFR>60 and CKD3+ groups three different features were selected for optimal separation as shown in Figure 6.8, although considerable overlap of the two groups was observed.

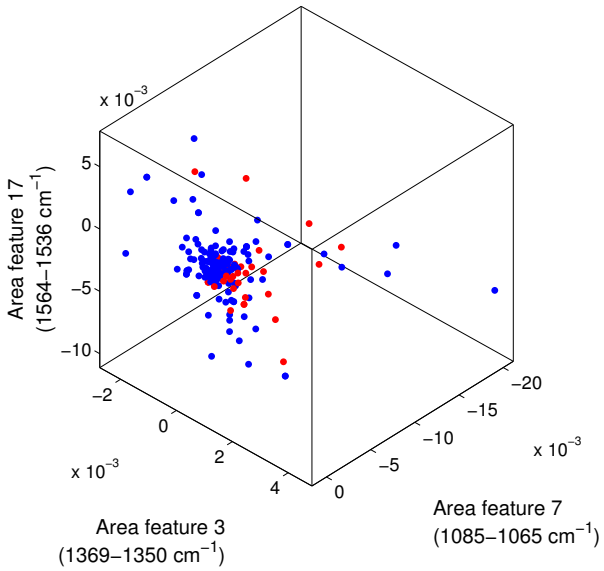


Figure 6.8: CKD3+ vs. the rest scatter plot. *Three-dimensional scatter plot comparing the areas of 3 spectral features found to be significantly different between the CKD3+ (red, abnormal) and eGFR>60 (blue, normal) groups.*

6.6.2 Patients with eGFR below 30 mL/min (CKD stage 4+)

Patients with eGFR values lower than 30 mL/min are classified as having stage 4 CKD (CKD4+), corresponding to severe kidney damage and kidney function of less than 30%. Urine spectra from these patients were compared with healthy urine spectra and to a second disease group containing spectra of urine samples from patients with eGFR >30 mL/min. The averaged second derivative spectra of these three groups can be seen in Figure 6.9. Table 6.4 details the p-values for each feature comparison and indicates which peaks are significantly different between the healthy and diseased samples at significance values of 0.05 and 0.01.

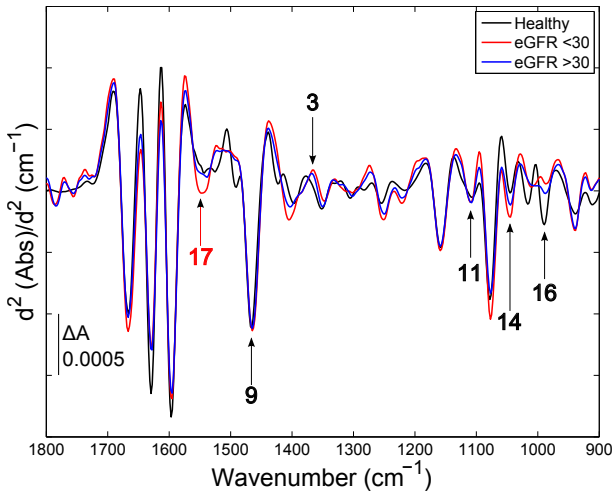


Figure 6.9: CKD4+ patient urine spectra. *Averaged second derivative urine spectra of all healthy (black), all CKD4+ (red, eGFR<30) and all eGFR>30 (blue) groups. Red arrows represent features found to be significantly different between both the healthy and CKD4+ groups and the eGFR>30 and CKD4+ groups at significance levels of at least 0.05. Black arrows represent additional features that showed significant differences between CKD4+ and healthy groups at significance levels of at least 0.05, but not between CKD4+ and eGFR>30 groups. All spectra were normalised to the area of their urea/creatinine band range at 1468 cm^{-1} between 1487 and 1436 cm^{-1}.*

Table 6.4: p values for comparison of 17 spectral features between healthy and CKD4+, and between CKD4+ and eGFR>30 samples. ** indicates features for which there was a difference between the two groups at a significance level of 0.05 and ** indicates differences at significance level of 0.01.*

Feature number	p-value CKD4+ vs. healthy	p-value CKD4+ vs. eGFR>30
1	0.81	0.01
2	0.81	0.33
3	0.98*	0.74
4	0.42	0.49
5	0.16	0.58
6	0.33	0.03
7	0.25	0.88
8	0.94	0.43
9	0.99**	0.37
10	0.05	0.32
11	0.98*	0.83
12	0.94	0.77
13	0.93	0.69
14	0.98*	0.13
15	0.87	0.05
16	0.99**	0.50
17	0.99**	0.98*

A 2D scatter plot comparing three features that gave the best separation of the healthy and CKD4+ groups is shown in Figure 6.10.

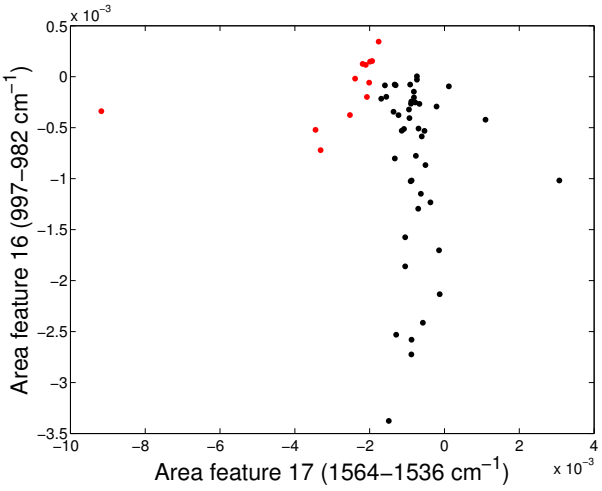


Figure 6.10: CKD4+ vs. healthy scatter plot. *Two-dimensional scatter plot comparing the areas of 2 spectral features found to be significantly different between the CKD4+ (red) and healthy (black) groups.*

Only one spectral band, feature 17, was found to be significantly different between the CKD4+ and eGFR>30 groups at a significance level of 0.05 but not at 0.01. A plot of the mean areas of feature 17 for each of the CKD4+, eGFR>30 and healthy groups, indicates that the area of feature 17 becomes more negative (corresponding to an increasing trough size) with increasing kidney damage (Figure 6.11).

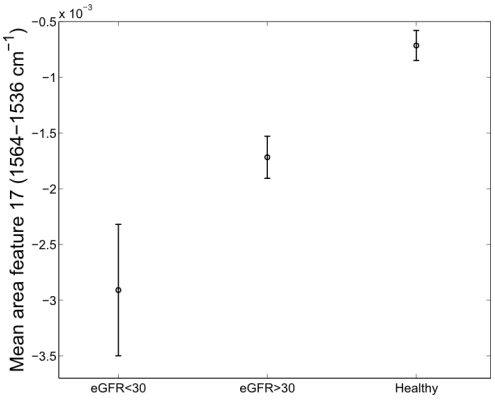


Figure 6.11: Comparison of amide II absorbance between diseased and healthy groups. *Plot of the mean areas of feature 17 (amide II band) for the CKD4+, eGFR>30 and healthy groups. Error bars represent standard error of mean.*

6.7 Discussion and conclusions

As well as an exploratory investigation of the diagnostic potential of urine FTIR spectra, this project aimed to establish the feasibility of installing an infrared machine in a hospital clinic to be run by a nurse. It was found that the staff nurses operating the machine were able to incorporate easily the measurement of urine FTIR spectra in their normal working day and record good quality, useable spectra. Two hundred and ninety-five spectra were recorded and 206 were taken forward for analysis. Of the 89 spectra not used in the analysis, most were rejected due to poor signal-to-noise. A small number were rejected because of erroneous peaks in the spectrum. In particular, spectra containing negative absorbance bands were removed as this indicates that the prism was contaminated when the background scans were recorded and that this contamination was subsequently removed when the prism was cleaned. Low signal to noise spectra generally corresponded to very dilute urine samples.

Examination of the individual spectra after pre-processing revealed that several biochemical components can be identified and that these displayed significant intensity differences between between spectra from different patients. Compounds that have particular diagnostic relevance include glucose and protein, which may be elevated in the urine of patients with diabetes or kidney damage. Glucose and other sugars absorb strongly in the 1000-1150 cm^{-1} region whilst protein has a distinctive amide II absorption band at around 1550 cm^{-1}. These features can be identified with confidence when they appear in urine spectra. Other spectral features were also identified that were not attributed to a specific biochemical, but may have diagnostic implications.

For all the disease groups studied there were clear differences in their urine spectra compared with the healthy group. However, it would be more useful clinically to be able to distinguish different forms of kidney injury and disease. Unfortunately, the sample numbers are currently too small to allow a robust analysis of individual kidney disease groups. Furthermore, many of the patients included in the study suffered from multiple kidney-related illnesses or other confounding conditions.

One factor that could be compared to IR features is glomerular filtration rate (GFR), which is used clinically as a measure of overall kidney function and can be a useful indicator of disease. In healthy young adults GFR is generally around 120 to 130 mL/min but a decline in GFR is considered part of the normal aging process. Excessive reduction in GFR is however used as an indicator of chronic kidney disease (CKD). CKD is defined as the condition in which kidney damage or a decrease in kidney function persists for 3 months or longer. CKD can have multiple causes and as such describes a

number of conditions in which kidney damage is sustained. When GFR falls below 60 mL/min there is an increased risk of complications associated with CKD as 50% or more of normal kidney function is lost. A GFR of 15 mL/min indicates kidney failure. Markers of CKD include proteinuria (protein in the urine), an albumin to creatinine ratio greater than 30 mg/g in spot urine, abnormal urine and serum biochemistries and abnormal urine sedimentation. CKD is staged according to the level of kidney function as defined by the GFR. GFR is difficult to measure directly and is therefore estimated from serum creatinine levels using, for example, the following Modification of Diet in Renal Disease (MDRD) equation:

$$GFR(mL/min\,per\,1.73m^2) = 186 \times (S_{Cr})^{-1.154} \times (Age)^{-0.203} \times (0.742\,if\ female) \times (1.210\,if\ African-American) \tag{6.1}$$

where S_{Cr} is serum creatinine concentration in mg/dL and age is in years [164]. An alternative equation, the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation, is also widely used and is reportedly more accurate at higher GFR values:

$$\begin{aligned} GFR &= 141 \times min(S_{Cr}/\kappa, 1)^{\alpha} \times max(S_{Cr}/\kappa, 1)^{-1.209} \\ &\quad \times 0.993^{Age} \times (1.018\,if\ female) \times (1.159\,if\ black) \end{aligned} \tag{6.2}$$

where S_{Cr} is serum creatinine, κ is 0.7 for females and 0.9 for males, α is -0.329 for females and -0.411 for males, min indicates the minimum S_{Cr}/κ or 1, and max indicates the maxS_{Cr}/κ or 1 [165]. However, CKD is underdiagnosed due to disagreement in the staging classification, lack of accuracy in eGFR calculations and lack of diagnostic tests. Furthermore, patients with 'normal' eGFR may still exhibit markers of kidney damage, such as proteinuria, which could go undetected. As a result preventative treatment cannot always be administered [162, 164–167].

Comparison of urine spectra from patients with abnormal eGFR with healthy patients showed that there were differences in several spectral features that allowed good separation of the two groups. A similar comparison of eGFR>60, CKD3+ and CKD4+ patient urine indicated that it might be possible to discriminate among patients based on the extent of kidney damage. This method could be used as a rapid screening tool to determine whether a patient has kidney damage and to indicate the extent of the damage, irrespective of the underlying cause. In addition, other diagnostic information,

including proteinuria, could be obtained without the need for further tests which could be useful for detecting early stage disease.

Many of the patients in this study were on medications for treatments of kidney-related diseases and other conditions. In some cases, this will reduce the symptoms of a disease and could lead to a patient presenting as 'normal'. For example, elevated urinary glucose in an untreated diabetic patient may return to a normal level on starting treatment. It may therefore not be possible to distinguish a true healthy patient from one whose disease is being corrected with medication. An additional consideration concerning medication is that some medicines will be excreted in the urine (as either native drug or as a derivative) and may lead to drug-associated spectral features. While this may complicate analyses of urine spectra, if a drug can be detected in urine it may be a useful measure of the patient's compliance to their treatment programme.

Other confounding variables may include age, sex and race related differences in the levels of some compounds in urine. The sample size is currently not large enough to reliably account for such differences. In addition and importantly, this study was unable to take into account the time of day of urine collection, which is likely to result in large variations in the urine composition because of factors such as diet, activity level and timing of medication administration. It would be informative to track a single patient with a known disease over a period of time of treatment, with collection and analyses of multiple urine samples at regular intervals, to determine whether any patterns of IR features and their responses to treatments can be detected.

This initial study on the feasibility of using a benchtop infrared spectrometer 'at the bedside' in a hospital clinic has shown that it has potential as a diagnostic tool. It was possible to detect a number of biochemical compounds in a single urine spectrum. Further data collection and analyses development will be required to establish to what extent this method can give more detailed diagnostic information for specific diseases.

Chapter 7

Detection of pancreatic cancer by ATR-FTIR spectroscopy

7.1 Introduction

Cancer is a progressive disease characterised by uncontrolled cell proliferation. This occurs when the tightly controlled mechanisms for regulating cell growth and replication fail and the tumour mass is able to induce angiogenesis and activate metastasis. Accumulation of epigenetic and genetic damage involving mainly oncogenes, tumour suppressor genes and DNA repair genes all contribute to the development of a full blown cancer. These processes are accompanied by biochemical changes in both tumour and neighbouring cells that can potentially be detected before any significant structural changes occur [168, 169].

Tumours range in aggressiveness from slow growing benign lesions to fast growing fully invasive and metastatic cancers, with 5-year survival rates varying accordingly. Pancreatic cancer is associated with a very poor 5-year survival rate of 5 %, and a 10-year survival rate of only 1 %. It is the fourth most common cause of cancer deaths in men and the fifth in women, and there has been little improvement in survival rates since the 1970s. The poor survival statistics are in large part a result of diagnosis only at the advanced metastatic stage of the disease, since there are no current screening programmes and symptoms tend to be general or absent in the early stages [170–172]. Although population screening for pancreatic cancer is not currently practised, some individuals have an increased risk of developing pancreatic cancer and can be monitored for disease even if symptomless. In particular, there are known genetic aberrations associated with pancreatic cancer and individuals with an affected first-degree relative are particularly at risk. Lifestyle factors such as diets high in fat, meat and nitrates, and medical conditions including diabetes mellitus and chronic pancreatitis also increase the risk of developing the disease [170–177].

Tumours of the pancreas are categorised based on the types of cells from which they originate; these may be cells of the exocrine pancreas, the endocrine pancreas or the lymphatic system of the pancreas. Ductal adenocarcinomas (ADCs), the most common type of pancreatic cancer, originate in the exocrine cells which are responsible for producing digestive juices. Across North America and parts of Europe and South America the incidence rate of pancreatic ductal ADC is approximately 10 in 100 000 per year, increasing to 200 in 100 000 per year for over-80-year-olds. Pancreatic neuroendocrine tumours (pNETs) originate in the endocrine cells where hormones such as insulin are produced. Ampulloma tumours arise at the boundary of the pancreatic duct, bile duct and duodenum. Although not technically a pancreatic tumour the ampulloma is treated in much the same way as an exocrine-type malignancy. The symptoms, treatments and

prognosis of exocrine and endocrine tumours differ (Cancer Research UK) [172, 177].

Staging of pancreatic tumours usually follows the widely used TNM system. The tumour is assessed for size and invasion of surrounding tissue (T), spread of cancer cells to regional lymph nodes (N) and presence or absence of distant metastases (M) in which the cancer has spread to distant lymph nodes or other tissues or organs. Hyperplasia, metaplasia and dysplasia may be observed in pre-invasive and early stage tumours although these are not well documented in the literature [174, 178]. The TNM stage can be summarised to give an overall stage or score from 0 to 3, where 0 represents benign tumours and 3 represents the most aggressive malignant tumours. Alternative staging scores based on the tumour's resectability may also be assigned [170].

Diagnosis and staging may be achieved by histological analysis of a biopsy section, or through other methods such as imaging. Computed tomography (CT) is the imaging method of choice, but MRI or ultrasound can be used as alternative or complementary techniques. If a diagnosis can be confirmed through imaging methods alone then a biopsy confirmation before tumour resection is not necessarily required. However, histological analysis of the resected tumour is usually performed [170, 173, 175, 179].

There is a need for more sensitive diagnostic procedures that can detect early stage disease and differentiate tumour types. Currently, the high false negative rate of pancreatic biopsy analysis has resulted in clinicians operating on patients without prior biopsy. In addition, different tumour types are associated with different prognoses so a more reliable biopsy analysis method would better inform the decision on treatment and whether or not to operate.

ATR-FTIR spectroscopy might provide a rapid and quantitative method for analysing pancreatic biopsies and has the potential to detect the subtle compositional changes that are associated with the early stages of tumour development. Whilst there is a rich literature describing the application of FTIR spectroscopy to the diagnosis of many other cancers, there are only a handful of groups who have attempted applying the technology to pancreatic cancer.

Kondepati et al. (2006) [180] studied DNA isolated from pancreatic adenocarcinoma samples for structural changes associated with carcinogenesis. Spectra were classified by HCA and peaks responsible for clustering were identified. Cluster analysis using the 1192 – 1059 cm^{-1} spectral region, which includes contributions from the deoxyribose-phosphate backbone of DNA, discriminated the tumour and normal DNA spectra with a sensitivity of 87 % and specificity of 80 %, where two tumour and three normal spectra were misclassified.

Another study sampled normal and cancerous tissue from the pancreas of a single patient for investigation by FTIR imaging [181]. Amide bands at 1659, 1547 and 1232 cm^{-1} were all reportedly reduced in cancerous samples when compared to normal tissue, along with other bands associated with protein. Decreases were also seen in the intensity of bands associated with lipid in cancerous tissue. However, there was no mention of spectra normalisation so these changes may be an artifact of sample thickness.

Clearly, further investigations are warranted of the FTIR characteristics of normal and tumorous pancreatic tissue samples, particularly to identify early changes in cells that could be used as an early detection method.

Pancreatic tissue resections were collected during major pancreatic surgery, in most cases samples were matched so that a sample of pancreatic tumour tissue and a sample of 'healthy' pancreatic tissue was taken from each patient. Therefore, all 'normal' samples were taken form patients with a tumour elsewhere in the pancreas. ATR-FTIR spectra of 43 fresh pairs of matched human pancreas tissue resections (43 tumour and 43 healthy) plus 3 unmatched tumour resections and 1 unmatched normal pancreas resection were recorded without any sample manipulation. Tumour staging information was also recorded and spectra of normal pancreas were compared to tumours of different stages in an attempt to identify the spectral features associated with pancreatic tumour progression. Table 7.1 summarises the number and stages of the pancreatic tissue resections received.

Table 7.1: Summary of numbers and stages of pancreatic tumour samples received.

Tumour type	No. samples	No. spectra
Stage 0 (benign)	6	12
Stage 1	10	20
Stage 2	13	26
Stage 3	17	34
Normal (non-tumour)	44	86

7.2 Summary of ATR-FTIR spectra collected of pancreatic resections

For detailed analyses the spectra were converted to their second derivatives to remove broad baseline artefacts, sharpen peaks and allow easier deconvolution of overlapping bands. The individual processed second derivative sample spectra can be

seen in Figure 7.1 and the averaged second derivative spectra of all tumour samples and of all normal pancreas samples are plotted in Figure 7.2, along with the computed difference between them.

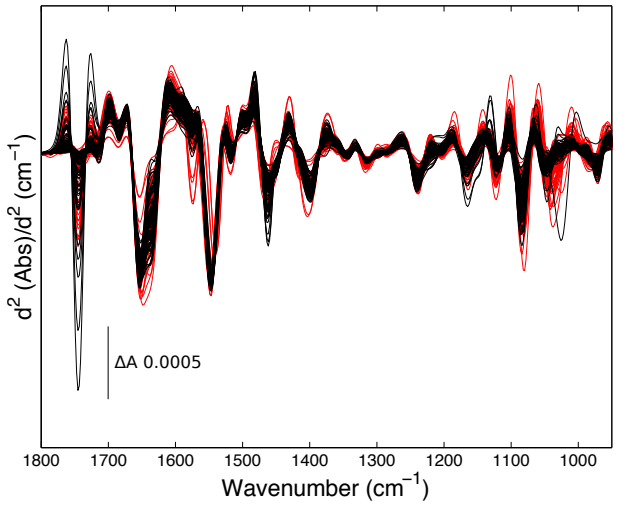


Figure 7.1: Individual second derivative spectra of normal and tumour pancreatic tissue. *Normal pancreas tissue spectra are shown in black and tumour sample spectra are shown in red. The 1800 – 950 cm^{-1} range of the fingerprint region only is shown. All spectra were normalised to the intensity of the amide II band (intensity at 1547 cm^{-1} minus intensity at 1527 cm^{-1}).*

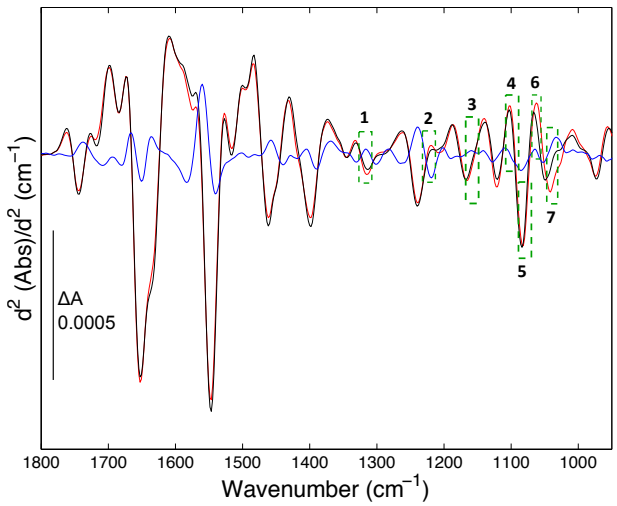


Figure 7.2: Averaged second derivative spectra of normal and tumour pancreatic tissue. *The black spectrum represents the mean of all normalised normal pancreas tissue spectra and the red spectrum is the mean of all normalised tumour sample spectra. The blue trace shows the difference spectrum. Green boxes represent the peaks chosen for integration and comparison between normal and tumour samples.*

7.3 Integration analysis of healthy and cancerous pancreas biopsy spectra

Seven potential peaks of interest/differences between the pancreas and tumour spectra were identified. The frequency limits for integrations were chosen based on a comparison of Figures 7.1 and 7.2. The chosen features are summarised in Table 7.2.

Table 7.2: Summary of potential features selected to distinguish between normal and tumour pancreas samples.

Feature Number	Wavenumber range (cm^{-1})
1	1329 - 1311
2	1232 - 1213
3	1169 - 1146
4	1105 - 1092
5	1088 - 1070
6	1066 -1053
7	1051 - 1028

The integrals of the areas of these features were calculated for each sample spectrum and were plotted against one another to generate 3D scatter plots. A combination of features 2, 3 and 7 were found to give the best separation of normal and tumour samples in a scatter plot of this kind (Figure 7.3).

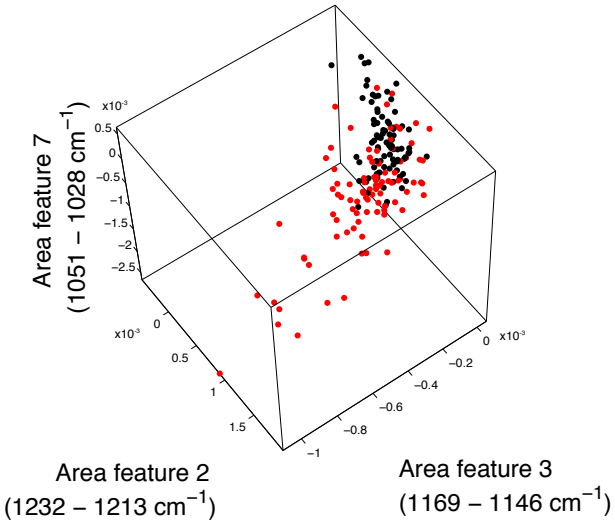


Figure 7.3: Scatter plot comparing normal pancreatic and tumour tissue. *Comparison of integrals of spectral features 2, 3 and 7 (Table 7.2). Each data point represents a single spectrum from a normal pancreas (black) or tumour (red) tissue sample.*

The tumour samples were scored/staged from 0 (benign) to 3 (most advanced stage) based on histological analysis under the light microscope of H&E stained tissue sections (this was performed externally by Dr. Jorge Paulino Pereira, Universidade Nova de Lisboa, Portugal). Scores were assigned based on the TNM staging system. The mean averages of the IR spectra of each of the tumour groups (0-3) were computed and compared to determine whether there were any spectral changes that could be correlated with tumour progression stage (Figure 7.4). Figure 7.5 is a replication of Figure 7.3 in which each point has been coloured according to its corresponding staging score.

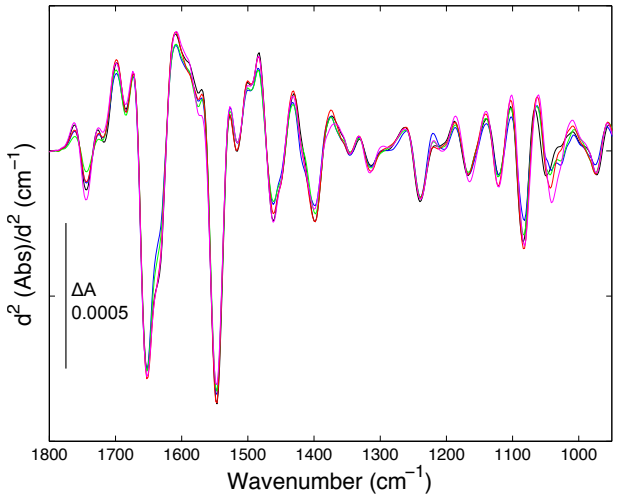


Figure 7.4: Averaged second derivative spectra of normal pancreatic tissue and tumour stages 0-3. *Second derivative spectra of normal pancreas (black) and stage 0 (blue), stage 1 (green), stage 2 (red) and stage 3 (magenta) tumour samples.*

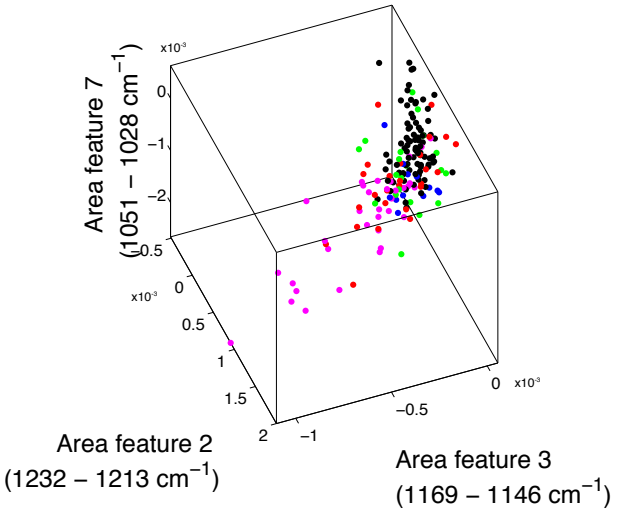


Figure 7.5: Scatter plot comparing normal pancreatic and tumour stages 0-3. *Comparison of integrals of spectral features 2, 3 and 7 (Table 7.2). Each data point represents a single spectrum from a normal pancreas (black) or a stage 0 (blue), 1 (green), 2 (red) or 3 (magenta) tumour tissue sample.*

Figures 7.4 and 7.5 indicate that there are several spectral transitions that may be correlated with different stages as the tumours develop although there is considerable overlap between samples. The clearest transition can be seen in the band at ~1050-1000 cm^{-1}, where there is both a shift to lower wavenumbers and an increased absorbance intensity (larger trough) from normal to stage 3 tumour. Other possible transitions occur at ~1310 cm^{-1} and at ~1120 cm^{-1}.

7.4 PLS-logistic regression model: staging of tumour tissue

Partial least squares (PLS) logistic regression of the IR spectral data was used in order to improve the separation of different stages, to make the analysis more statistically robust and to provide quantitative results. PLS was used to minimise the chance of overfitting by reducing the dimensionality of the data whilst maintaining the relationship between the spectral pattern and the disease groups. Logistic regression is a common classification method to use on categorical data, such as this pancreatic spectra dataset. After optimisation, it was found that three PLS-logistic regression models used in series gave the best separation of the data. Figure 7.6 summarises the proposed diagnostic pipeline.

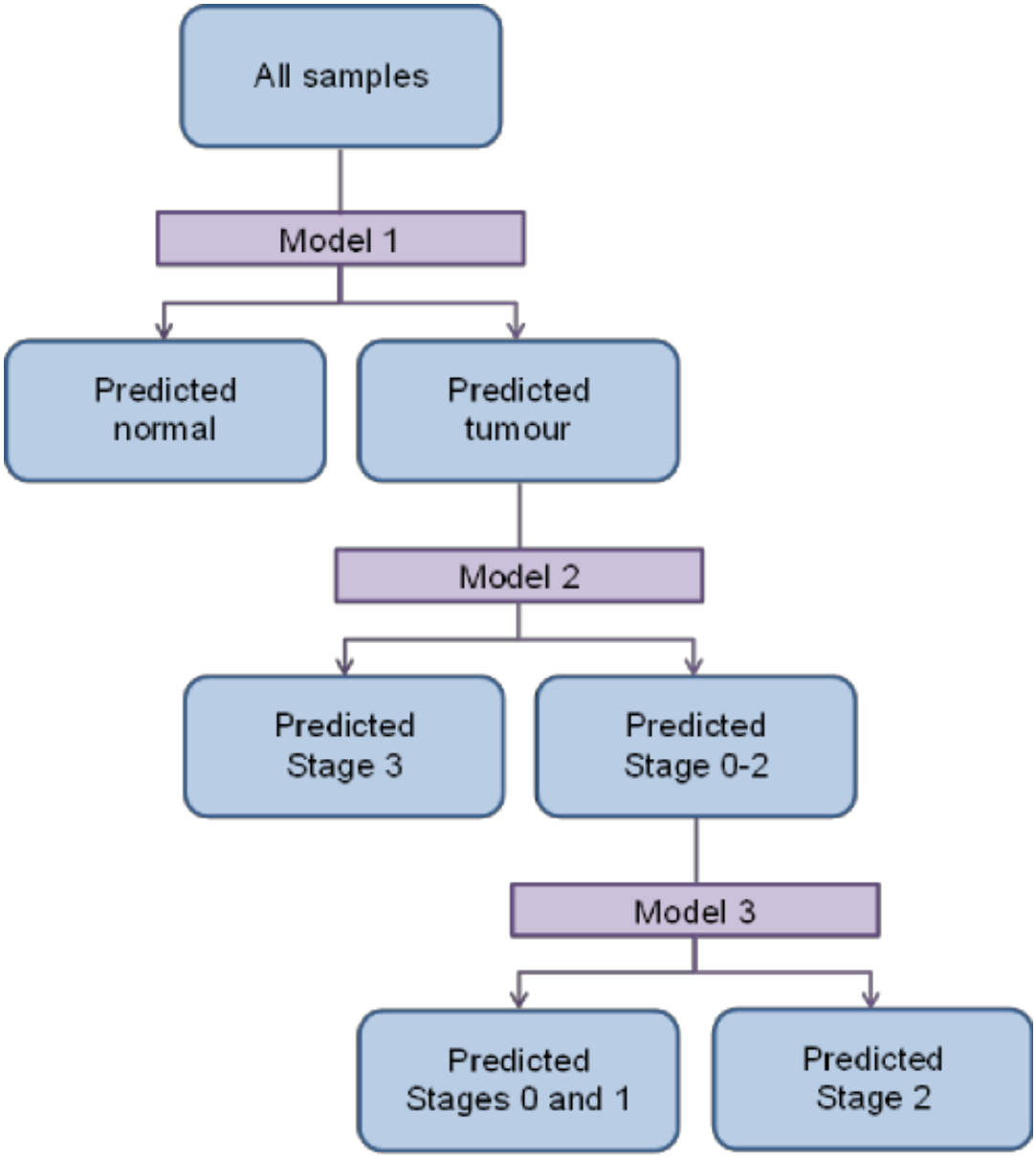


Figure 7.6: Proposed diagnostic pipeline for classifying FTIR spectra from pancreatic cancer patients according to tumour stage using PLS logistic regression.

Model 1: separating all normal spectra from all tumour spectra. The 1175 - 1025 cm^{-1} regions of the second derivative spectra were extracted and mean centred. Five PLS latent variables (LVs) were chosen for the logistic regression because this gave a good compromise between specificity and sensitivity of separation into the two classes. A cost of 1.5 was applied for incorrectly assigning a tumour spectrum as normal (correct classification of normal and tumour had no cost).

Model 2: separating stage 3 tumour spectra from all other tumour spectra. The 1450 - 1050 cm^{-1} regions of the second derivative spectra were extracted and

mean centred. Six PLS LVs were chosen for the logistic regression. A cost of 1.25 was applied for incorrectly assigning a stage 3 tumour spectrum as any other stage tumour (correct classification had no cost).

Model 3: separating stage 0 and 1 tumour spectra from stage 2 tumour spectra. The 1150 - 1025 cm^{-1} regions of the second derivative spectra were extracted and mean centred. Five PLS LVs were chosen for the logistic regression.

In all cases, leave-one-patient-out cross validation was used to assess the predictive ability of the PLS logistic regression model.

7.4.1 Performance of model 1: separating all normal spectra from all tumour spectra

Before assessing the performance of the model it was important to confirm that the largest features in the loadings plots (i.e. those features with the biggest contribution) corresponded to true signal variations rather than differences in noise. Examination of the loadings on the first five LVs of model 1 (Figure 7.7) confirmed that they are likely to represent true signal differences between normal and tumour sample spectra.

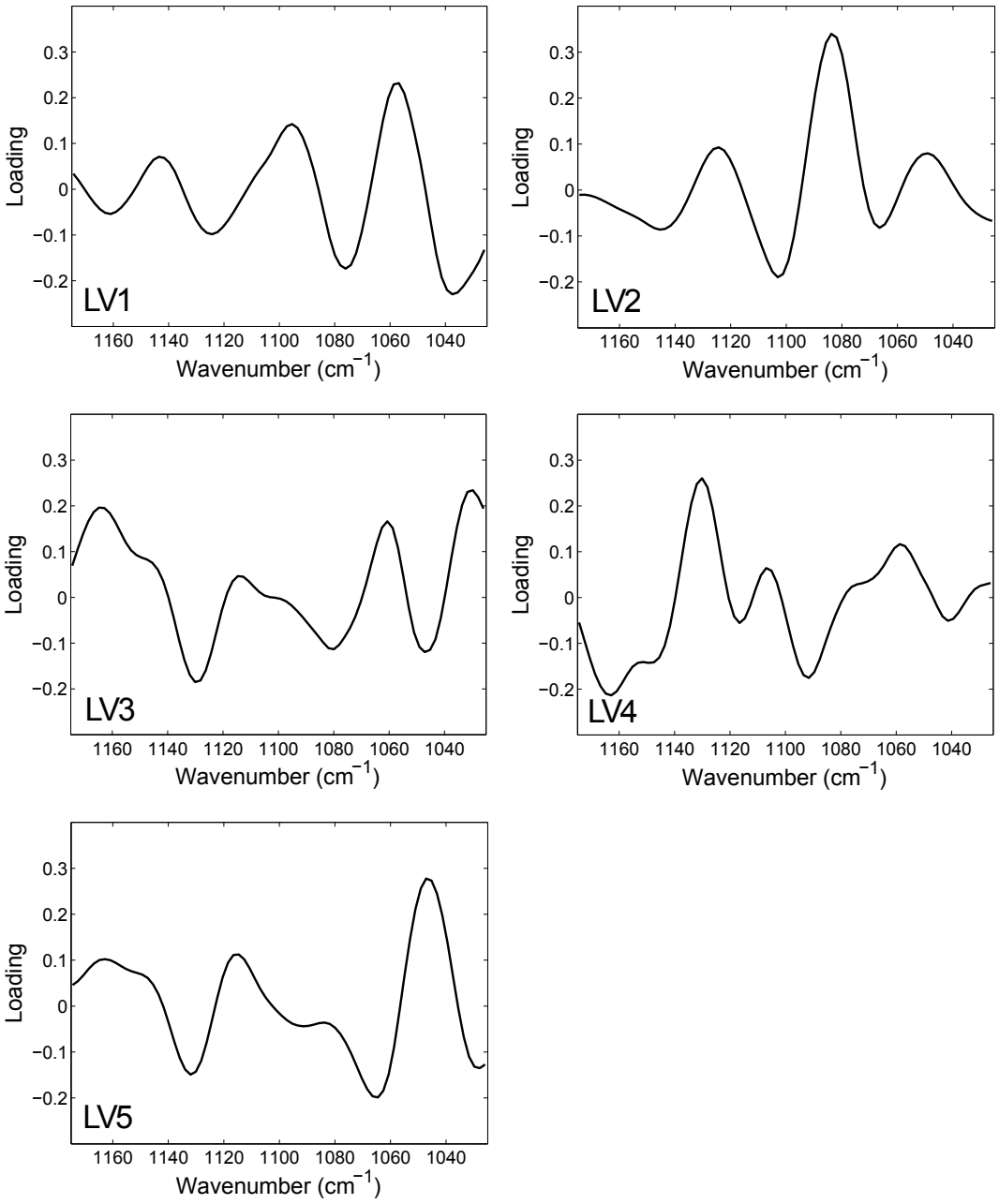


Figure 7.7: Loadings on the first five latent variables for PLS logistic regression model 1.

Figure 7.8 shows the scatter plot of scores of LVs 1, 2 and 4 for each spectrum.

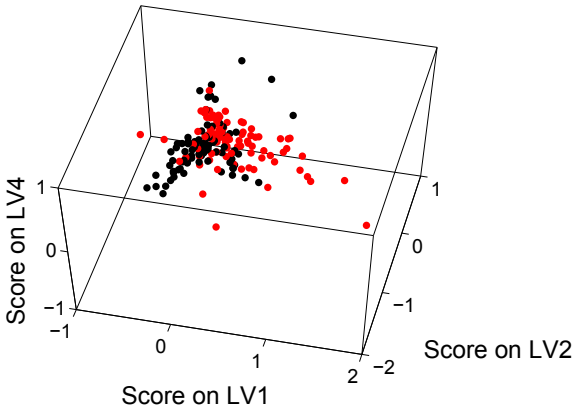


Figure 7.8: 3D scatter plot for model 1. *Scores plot of LVs 1, 2 and 4 for normal pancreas (black) and tumour (red) tissue samples.*

The scores on the LVs for each spectrum were used to build the logistic regression model. Probability scores were computed for each spectrum that describe the likelihood of that spectrum belonging to the either the normal or the tumour classes. The final assignment of the spectrum to either class was achieved by selecting the class associated with the highest probability. Since this was a two-class model this is equivalent to setting a class threshold of 0.5 where a value closer to 1 (above 0.5) indicates membership to the given class. Each spectrum was assigned by choosing the most probable class, meaning every spectrum was classified into one or other class and none were unassigned.

Table 7.3 shows the confusion table for the Normal vs Tumour PLS logistic regression model (model 1), where values on the main diagonal (top left to bottom right) show the number of correctly assigned spectra and the off-diagonal values are the numbers of incorrect assignments. Calculated sensitivities and specificities for predicting spectra as normal or tumour are listed in Table 7.4. Sensitivity is a measure of the model's ability to identify all spectra that belong to the class in question (i.e. the extent that false negatives are minimised). Conversely, specificity is measure of the model's ability to exclude all other classes from the class in question (i.e. the extent that false positives are minimised). The overall accuracy of the model was 84%. To maximise the sensitivity of the model for detecting tumour spectra a cost was applied for incorrect classification of tumour spectra as normal. Consequently, the specificity of tumour detection is decreased and an increased number of normal spectra were classified as tumour. If an equal cost for incorrect classification was applied the sensitivity of detecting normal spectra increased to 81% but the sensitivity of detecting tumour spectra fell to 82%.

Table 7.3: Confusion table for model 1. *Confusion table showing number of normal and tumour spectra correctly predicted after cross validation.*

		Predicted class	
		Normal	**Tumour**
Actual class	**Normal**	66	20
	Tumour	9	83

Table 7.4: Sensitivities and specificities for model 1. *Sensitivities and specificities for predicting normal and tumour spectra after cross validation.*

	Sensitivity (%)	**Specificity (%)**
Normal	77	90
Tumour	90	77

7.4.2 Performance of model 2: separating stage 3 tumour spectra from stages 0-2 tumour spectra

A second PLS logistic regression model was built to separate tumour spectra according to stage. Since the stage 3 tumour group was bigger than any of the stage 0, 1 or 2 tumour groups a model was built to first separate the stage 3 tumour spectra from the rest of the tumour spectra. Examination of the loadings on the first 6 LVs (Figure 7.9) confirmed that they are likely to represent true signal differences between normal and tumour sample spectra.

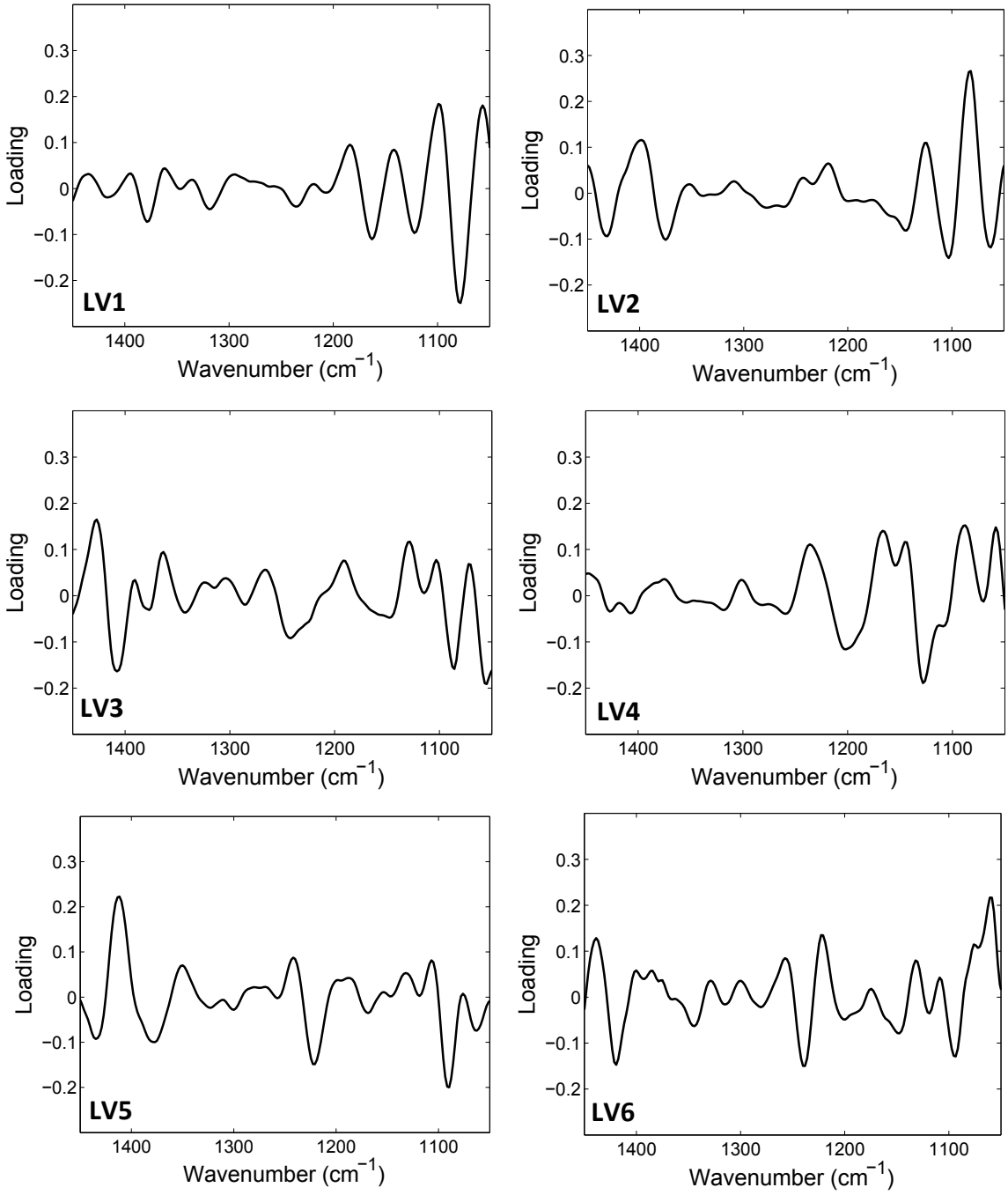


Figure 7.9: Loadings on the first 6 LVs for PLS logistic regression model 2.

Figure 7.10 shows a scatter plot of the scores of LVs 1, 2 and 5 for each spectrum.

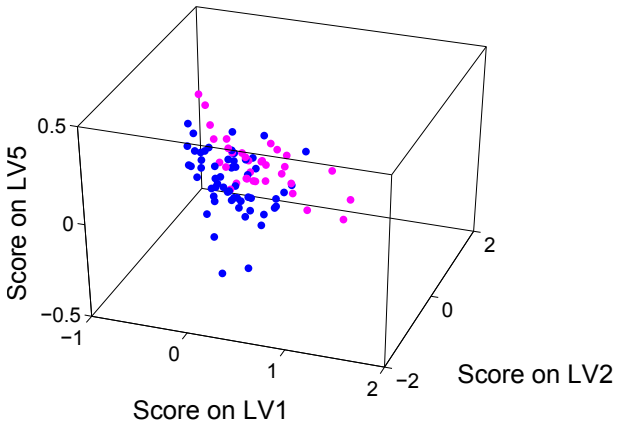


Figure 7.10: 3D scatter plot for model 2. *Scores plot of LVs 1, 2 and 5 for stage 0 - 2 tumours (blue) and stage 3 tumours (magenta).*

Table 7.5 shows the confusion table for model 2 (stage 0 - 2 tumours *versus* stage 3 tumours) and calculated sensitivities and specificities for predicting spectra as stage 0 - 2 or stage 3 are listed in Table 7.6. The overall accuracy of this model was 84%.

Table 7.5: Confusion table for model 2. *Confusion table showing number of stage 0 - 2 and stage 3 tumour spectra correctly predicted after cross validation.*

	Predicted class		
		Stage 0 - 2	**Stage 3**
Actual class	**Stage 0 - 2**	50	8
	Stage 3	7	27

Table 7.6: Sensitivities and specificities for model 2. *Sensitivities and specificities for predicting stage 0 - 2 and stage 3 tumour spectra after cross validation.*

	Sensitivity (%)	**Specificity (%)**
Stage 0 - 2	86	79
Stage 3	79	86

7.4.3 Performance of model 3.: separating stage 2 tumour spectra from stages 0 & 1 tumour spectra

Since early-stage detection of pancreatic cancer is key to increasing survival rate, a model was built to try to further distinguish early-stage tumour spectra. Limited availability of early stage tumours meant that there were relatively few samples in each

of the stage 0, 1 and 2 classes, therefore stage 0 and stage 1 tumours were grouped into a single class. Five PLS LVs were used in the PLS regression model and the loadings on these LVs are shown in Figure 7.11 to demonstrate that they represent true signal differences.

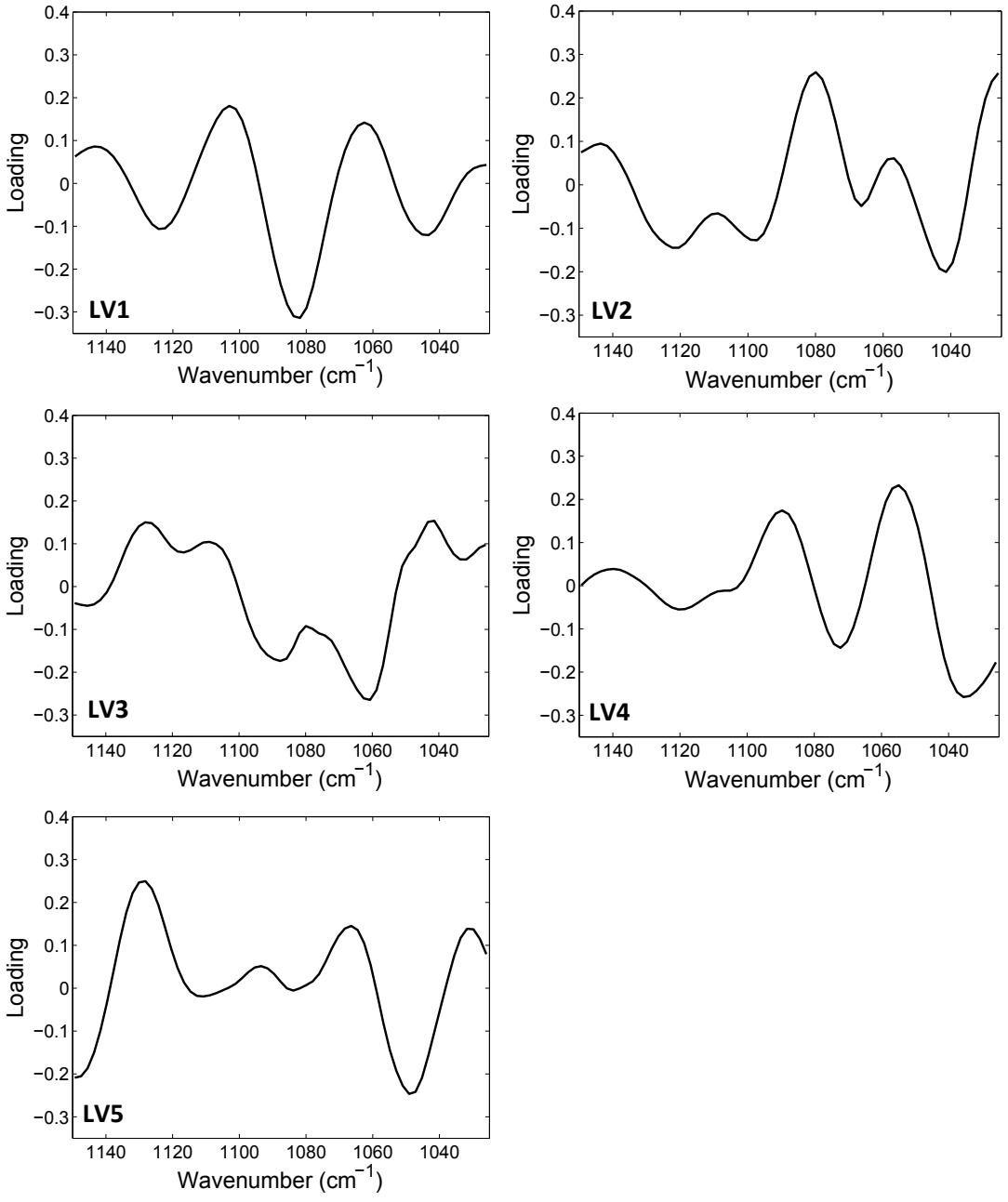


Figure 7.11: Loadings on the first 5 LVs for PLS logistic regression model 3.

Figure 7.12 shows a scatter plot of the scores of LVs 3, 4 and 5 for each spectrum.

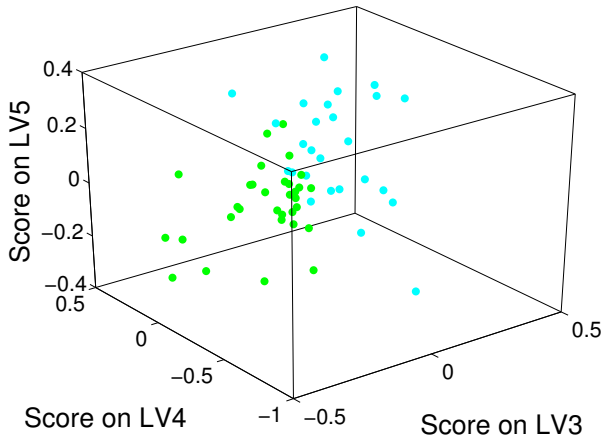


Figure 7.12: Scores plot for model 3.. *Scores plot of LVs 3, 4 and 5 for stage 0 and 1 tumours (green) and stage 2 tumours (cyan).*

Table 7.7 shows the confusion table for model 3 (stages 0 and 1 tumours *versus* stage 2 tumours) and calculated sensitivities and specificities for predicting spectra as stages 0 and 1 or stage 3 are listed in Table 7.8. The overall accuracy of this model was 64%.

Table 7.7: Confusion table for model 3. *Confusion table showing number of stages 0 and 1 and stage 2 tumour spectra correctly predicted after cross validation.*

	Predicted class		
		Stage 0 and 1	**Stage 2**
Actual class	**Stage 0 and 1**	23	9
	Stage 2	12	14

Table 7.8: Sensitivities and specificities for model 3. *Sensitivities and specificities for predicting stages 0 and 1 and stage 2 tumour spectra after cross validation.*

	Sensitivity (%)	**Specificity (%)**
Stage 0 and 1	72	54
Stage 2	54	72

7.5 Correlation of pancreas biopsy spectra with model compound spectra

In general it is difficult to assign peaks in tissue spectra to specific biochemical components since the spectra are complex with many overlapping bands. More confident assignments can be made by correlating multiple peaks with model compound spectra, or by use of additional complementary biochemical tests. However, comparison to model compound spectra can give an indication of the types of compounds that might be contributing to the sample spectrum or to the differences between normal and diseased sample spectra.

In order to support continued proliferation, tumour cells have been reported to alter their metabolic pathways to generate the energy required for growth [168]. Changes in the content of carbohydrates, particularly glycogen, in tumour cells compared to normal cells have been reported by many groups who have used FTIR spectroscopy for cancer diagnostics [18,35,55,56,65,168]. Potential carbohydrate biomarkers for pancreatic cancer have also been proposed. For example, carbohydrate antigen 19-9 (CA-19-9) is an FDA-approved biomarker for pancreatic cancer, and the levels of other carbohydrates including CA-50 and CA-242 have been shown to be altered in pancreatic cancer. It is unlikely that FTIR spectroscopy would be able to detect changes in these specific carbohydrates but it may be able to report on global changes in carbohydrate content [182,183].

Figure 7.13 shows how the pancreatic tissue spectra compare to model spectra of glycogen and glucose. These sugars absorb strongly in the 1200 - 1000 cm^{-1} region, which is also the region associated with the biggest differences between the normal and tumour pancreatic tissue spectra, where there is an overall slight increase in absorbance intensity going from normal tissue to stage 3 tumours. Differences in carbohydrate content may contribute these spectral differences.

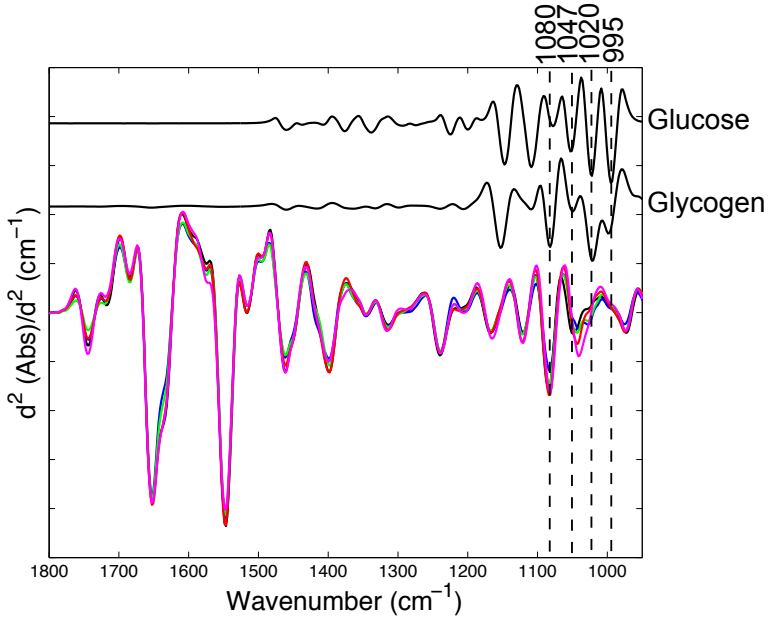


Figure 7.13: Comparison of pancreatic tissue spectra with model sugar spectra. *Model sugars shown are glycogen and glucose. Dashed lines indicated possible correlated peaks between the pancreatic tissue spectra and model sugar spectra. The overlaid spectra at the bottom of the plot represent normal pancreas (black) and stage 0 (blue), stage 1 (green), stage 2 (red) and stage 3 (magenta) tumour samples.*

Changes in RNA and/or DNA content of tumour tissue compared to normal tissue have been reported in FTIR spectroscopy studies of cancers [31,35,55,88], including in a study looking at pancreatic cancer [180]. These differences may come about through changes in chromosome ploidy of tumour cells and dysregulation of gene expression [184–186]. The model spectra of RNA and DNA (Figure 7.14) show that these compounds have significant absorbance bands between 1200 and 1000 cm^{-1} so may also contribute to the differences between normal and tumour pancreatic tissue spectra in this region.

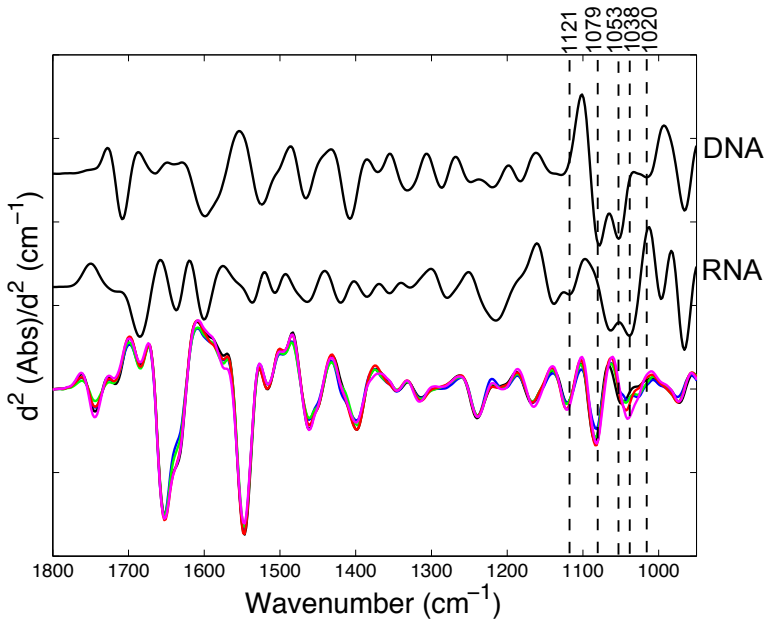


Figure 7.14: Comparison of pancreatic tissue spectra with model RNA and DNA spectra. *Dashed lines indicated possible correlated peaks between the pancreatic tissue spectra and model RNA and DNA spectra. The overlaid spectra at the bottom of the plot represent normal pancreas (black) and stage 0 (blue), stage 1 (green), stage 2 (red) and stage 3 (magenta) tumour samples.*

Although FTIR spectroscopy cannot generally distinguish between different types of protein, collagen and mucin are two proteins with unique compositions that produce some distinctive IR signals and changes in their cellular content have been correlated with pancreatic cancer. Mucin is a highly glycosylated protein that is produced by epithelial cells and is the main constituent of mucus. Membrane bound mucins also exist. Mucins are present in normal pancreatic tissue, but are also thought to be involved in tumour development due to their roles in cell signalling, differentiation and proliferation. Indeed, changes in the expression pattern of mucins during pancreatic tumour progression have been identified [182, 187, 188]. Collagen is an abundant extracellular protein found in connective tissue. One study found increased expression of type IV collagen in pancreatic tumour tissue and high levels of circulating type IV collagen in patients with the disease [189].

Figure 7.15 compares the pancreatic tissue spectra with collagen and mucin model spectra (with albumin included as a reference for general protein). Multiple bands can be correlated between the protein and sample spectra, particularly in between 1200 and 1000 cm^{-1} but the spectra are complex and specific assignment is not possible.

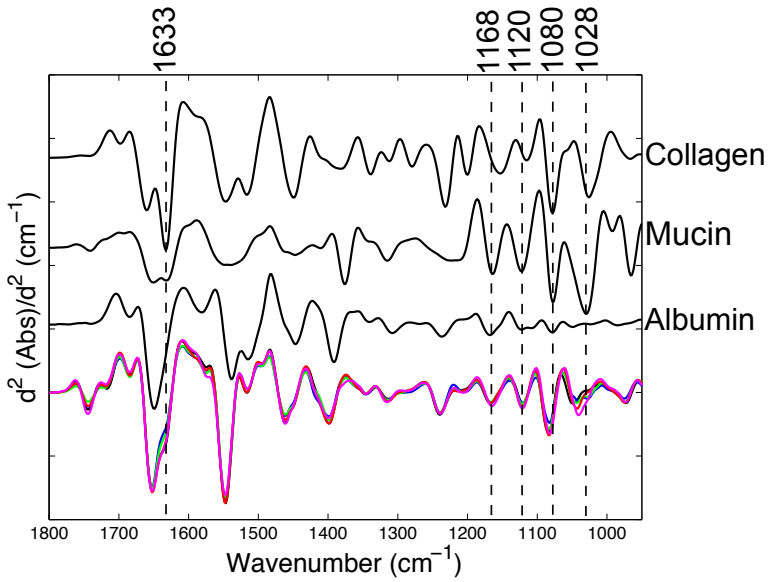


Figure 7.15: Comparison of pancreatic tissue spectra with model protein spectra. *Model proteins shown are albumin, mucin and collagen. Dashed lines indicated possible correlated peaks between the pancreatic tissue spectra and model protein spectra. The overlaid spectra at the bottom of the plot represent normal pancreas (black) and stage 0 (blue), stage 1 (green), stage 2 (red) and stage 3 (magenta) tumour samples.*

7.6 Discussion and conclusions

The individual normal and tumour tissue spectra and their respective averaged spectra show that there are several spectral differences between the normal and tumour tissue. Some of these differences are subtle (e.g. features 3, 4 and 5) whilst others are more pronounced (e.g. features 1, 6 and 7). One of the normal sample spectra was considerably different to the others and may represent an outlier. Examination of the individual sample spectra in Figure 7.1 reveals that the tumour samples exhibit more variation, particularly between 1200 and 1000 cm^{-1}, than the normal samples. This may be the result of increased heterogeneity in tumorous tissue compared to healthy tissue, or may arise from different stages of malignancy in the tumour samples.

The scatter plot in Figure 7.3 showed that, although some separation of the normal and tumour sample spectra could be achieved based on the integration of three features of interest, there was considerable overlap. PLS logistic regression gave better separation of the normal and tumour data (model 1) based on the scores on the LVs 1, 2 and 4 for each spectrum. Two clear clusters, representing the normal samples and the tumour samples, can be seen in Figure 7.8, with a smaller region of overlap than was observed with the peak integration method. Cross-validation allowed the predictive ability

of the PLSDA model to be assessed. Class assignments predicted by the model also indicated good separation of the normal and tumour spectra with the overall accuracy of the model being 84% with 90% sensitivity for detecting tumour spectra. The sensitivity for detection of tumour spectra was maximised by applying a cost to incorrectly classifying tumour spectra as normal (false negatives) during the model construction. This slightly reduced the specificity of detection of tumour spectra but minimises the chances of missing a tumour in a patient.

Model 1 was built using the 1175-1025 cm^{-1} region of the IR spectra of the pancreatic tissue sample. Kondepati et al. (2006) [190] used a similar region (1192-1059 cm^{-1}) to separate DNA from normal and tumour pancreatic tissue, suggesting that changes in the DNA or RNA content of the pancreatic tumour samples studied here may have contributed to the spectral differences. The model spectra of DNA and RNA (Figure 7.14) show that these compounds do have strong absorbance signals in this region, but these could not be distinguished from the signals of the other model compounds that absorb in this region.

It is important to note that the tumour spectra used to build model 1 included a range of tumours at different stages from 0 (benign) to 3 (most advanced). Since cancer is a progressive disease a region of overlap between normal and tumour might be expected in any diagnostic model, representing a transition from normal pancreas through early tumour stages to advanced stage tumour.

Since the ability to differentiate intermediate tumour stages is more clinically relevant than being able to separate normal and tumour samples (as the latter is already relatively straightforward using existing methods) an attempt was made to stage the tumour spectra using PLS logistic regression.

It was not possible to separate, with good accuracy, the sample spectra into normal and the four tumour stages using a single multinomial model. Therefore, a series of models was built to provide a diagnostic pipeline (Figure 7.6) in which all tumour spectra could be separated into stages 0 - 2 and stage 3 (model 2), and subsequently stage 0 - 2 tumour spectra could be separated into stages 0 plus 1 and stage 2 (model 3).

Model 2 performed with 84% overall accuracy, 86 % sensitivity for detecting stage 0 - 2 spectra and 79% sensitivity for detecting stage 3 spectra. Model 3 performed with 64% overall accuracy and 72% sensitivity for detecting stage 0 and 1 spectra. However the sensitivity for detecting stage 2 spectra using model 3 was only 54%.

The performance of all the models, but particularly those aimed at separating the tumour samples into stages, is compromised by the limited number of patient sam-

ples in each class. Expanding the dataset, to both increase total sample numbers and increase the proportion of early stage tumour samples, would likely further increase the accuracies of the models. Indeed, to reach statistically significant accuracy that meets clinical guidelines a much larger study would be required.

The models in the diagnostic pipeline described are designed in series so that the output of model 1 (those spectra predicted to be tumour) would pass into model 2 and, in turn, the output of model 2 (those spectra predicted to be stage 0 - 2 tumours) would pass into model 3. Since none of the models are 100% specific the error of one model is carried on to subsequent models. This means that the realised accuracies of models 2 and 3 are lower than the values quoted, as the output of the preceding model contains some incorrectly assigned spectra (false positives). For example, the model 1 predicts that 103 of the 178 spectra are tumour; of these, 83 are tumour and 20 are normal. However, all 103 are passed to model 2, which does not have the capacity to predict normal so must classify those 20 normal spectra as either stages 0-2 or stage 3. The maximum performance of model 2 is therefore 84% of the 84% accuracy of model 1.

Comparison of the pancreatic tissue spectra to a set of model compound spectra gives an indication of the kinds of compounds that might contribute to the sample spectra and to the differences between normal and tumour samples. The spectral region showing the most obvious differences between tumour stages and types is between 1200 and 1000 cm^{-1}. As demonstrated, the model spectra of mucin, glycogen, glucose, DNA and RNA all have strong absorbance bands in this region, as do other types of compounds, such as lipids. Other studies have also identified changes in these types of compounds in relation to the development of pancreatic cancer [180, 182, 187–189]. However, the 1200-1000 cm^{-1} region of the IR spectrum, identified here as being potentially significant in diagnosis, is complex and without further biochemical tests it is not possible to attribute the spectral changes to any specific compound(s).

Chapter 8

Concluding remarks and future prospects

This thesis concerns the application of ATR-FTIR spectroscopy to the diagnosis of kidney diseases and pancreatic cancer, as well as the influence of hydration state on IR spectra which has proved important when analysing dried biological samples. A step has also been taken towards translating IR technology from the research laboratory to the hospital clinic.

FTIR spectroscopy and its use in medical diagnostics is a fast growing field that has been applied to a wide range of diseases and samples types. In many ways, FTIR spectroscopy is an ideal technique for analysing biological samples because it is quantitative, sensitive to subtle compositional changes, and can provide a lot of information in a single spectrum. Furthermore, with advancing technology, in terms of both hardware and software, good quality IR spectra can be recorded and processed without detailed technical knowledge and with minimal training. However, IR spectra are subject to many types of interferences and artefacts that, if not adequately identified and dealt with, can lead to misinterpretation of spectra.

To illustrate this point, the influence of hydration state on the IR spectra of urinary components has been described. Dramatic changes were observed in the IR spectrum of urea with only a small change in humidity. Throughout the FTIR spectroscopy in medical diagnostics literature there are examples of dried sample analysis [73, 74, 83, 90, 91], but there is little indication that rehydration levels were controlled or considered. More generally, results of FTIR analyses are often presented without sufficient supporting spectra so that it is not possible to critically evaluate the validity of the work. This is a particular problem when complex statistical methods have been applied to data, as these methods are not capable of distinguishing real signal differences from spectral artefacts. It is therefore necessary that each step of spectral analysis is carefully monitored by someone who understands the complexity of the data and limitations of the technology. Ultimately, although IR spectroscopy is quantitative, a value judgement is always required at some stage to decide which spectral features and changes are biologically relevant.

The main focus of clinical FTIR spectroscopy is to provide new and improved methods of diagnosing diseases that are currently difficult or expensive to diagnose accurately using existing methods. Disease diagnosis using FTIR spectroscopy may be achieved by analysing a sample spectrum for specific biomarkers that have distinctive IR signals, or by looking for global pattern changes in spectra that can be correlated to disease but which do not necessarily have known origins.

Urine is a relatively simple mixture of known composition and it is possible to

identify and quantitate some of its constituents from its IR spectrum, and to identify the specific compounds that differ between two samples. Kidney stone diseases represent a good example of how FTIR spectroscopy can be used to detect specific disease biomarkers. Here, it has been demonstrated that the stone disease, cystinuria, can be diagnosed from ATR-FTIR spectra of the dried insoluble fractions of urine samples. It is likely that similar protocols could be used to detect any other insoluble stone component, provided it is IR active and present at a concentration sufficient to meet the sensitivity limit of the spectrometer. This was demonstrated for detection of DHA in the diagnosis of stone disease, APRT-deficiency, although limited samples were available. Kidney stone diseases are often difficult to diagnose at the early pre-stone stage, and distinguishing between stones types is also problematic , but with ATR-FTIR spectroscopy a single sample spectrum could be simultaneously screened for a wide range of kidney stone types.

One of the difficulties encountered in the analysis of urine samples is that components at high concentrations, such as urea, dominate spectra. If the component of interest is insoluble, as in the case of kidney stone diseases, centrifuging the samples is an effective way to remove the high concentration soluble components. However, for detection of some low concentration soluble biomarkers, further development may be required to maximise their IR signals. For example, an additional sample processing step could be used to separate soluble components, perhaps using the laminar flow microfluidic device [83] described in the introduction. Alternatively, an additional data processing step, such as subtracting model spectra of high concentration components from the sample spectrum, taking into account the hydration state as described in results Chapter 4, could be used to reveal underlying absorbance bands.

The laboratory-grade spectrometers that were used in the development of the cystinuria diagnostic protocol, and in much of the literature, require liquid nitrogen and are too large, expensive and complex to realistically be used routinely in hospital clinics, or even in pathology laboratories. In addition, the method of separating the insoluble material and drying it on the prism is relatively slow. If FTIR spectroscopy is to be used clinically for this purpose, research must move towards use of devices that can be installed in hospitals, and protocols must be simplified so that they can be carried out by a nurse or other staff member without major disruption to the clinical workflow. Sample processing must also be made high-throughput, and potentially automated, so that multiple samples can be processed and dried simultaneously and reproducibly.

The portable benchtop spectrometers that have recently been developed for field work, primarily in forensics and chemistry, are small and stable enough for in-

clinic use. One such machine was installed at the Royal Free Hospital and operated by staff nurses who were able to produce good quality data. Useful diagnostic information could be obtained from unprocessed wet urine samples that gave an indication of kidney function. However, it was not possible to identify the causes of loss of kidney function, meaning that additional tests are currently required for more specific diagnosis. With further research into the influence of interfering factors, such as age, ethnicity, diet and confounding diseases, it may be possible to extract more specific diagnostic information from such spectra. However, unless a disease is associated with a urinary biomarker that has a unique IR signal it is likely that this technology will be more useful as a screening tool rather than a standalone diagnostic.

The protocol for recording and analysing urine spectra on the benchtop spectrometer in the clinic was made as simple and automated as possible for proof of concept purposes. However, with some optimisation it should be possible to adapt the protocols developed for detecting cystine and other stone compounds so that they could be performed in the clinic for rapid kidney stone disease screening. Similarly, where other groups have used ATR-FTIR spectroscopy for disease diagnosis, using blood products, urine or other biofluids [73–78, 80], translation to in-clinic analysis should be possible.

In contrast to the relatively straightforward analyses of urine composition, analyses of pancreas tissue resections, described in results Chapter 7, is an example of how statistical methods can be used to identify spectral patterns, rather than specific biochemicals, associated with disease. It is not usually possible to assign bands with high confidence in tissue IR spectra as the spectra are complex with many overlapping bands, and different tissue constituents often exhibit very similar spectral signatures (e.g. different types of protein). It is generally only possible to make speculative assignments based on the spectral region in which an absorbance band falls, with reference to known biological processes that have been shown to be relevant to the disease process by complementary biochemical tests. Using PLS logistic regression to extract spectral patterns from the pre-processed ATR-FTIR spectra of pancreatic tissue resections, it was possible to discriminate between normal and tumour tissue, and to a lesser extent between early and late tumour stages. One of the biggest limitations of this study was the lack of availability of early-stage tumour samples, resulting in diagnostic models that may not accurately represent all tumours of a given stage. Furthermore, different cell types and tumour types may generate different absorbance patterns. Ideally, samples should first be classified by cell/tumour type before being further divided into disease stages.

There has been increased interest in the use of FTIR microscopy for cancer di-

agnostics, as well as for a number of other biological and diagnostic applications [14–19, 21, 27]. The popularity of this technique is due to the high spatial resolution that can be obtained and because it offers a visual means, comparable to light microscopy, of analysing a sample whilst obtaining a huge amount of biochemical information. However, although this technology can provide a unique insight in to cancer development, it is unlikely that it will be suitable for clinical use in the near future due to its cost, complexity and because of the major disruption this would cause to existing practice. For this reason, the focus here has remained on ATR-FTIR spectroscopy applications which, as already discussed, can be achieved with portable benchtop machines that could be used point-of-care.

For all of the applications of FTIR spectroscopy in medical diagnostics described, large scale clinical studies using protocols and hardware suitable for hospital installation are now required to prove the statistical significance and efficacy of the diagnostic methods and the potential benefits over existing methods.

Bibliography

[1] Siebert F and Hildebrandt P (2008). *Vibrational Spectroscopy in Life Science*. Wiley-VCH Verlag GmbH & co. KGaA, Weinheim, FRG.

[2] Wartewig S (2006). Basic principles of vibrational spectroscopy. In *IR and Raman Spectroscopy*, pages 27–33, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG.

[3] Goormaghtigh E, Raussens V and Ruysschaert JM (1999). Attenuated total reflection infrared spectroscopy of proteins and lipids in biological membranes. *Biochimica et Biophysica Acta* **1422**: 105–185.

[4] Movasaghi Z, Rehman S and Rehman Iu (2008). Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Applied Spectroscopy Reviews* **43**: 134–179.

[5] Carter EA, Tam KK, Armstrong RS and Lay PA (2009). Vibrational spectroscopic mapping and imaging of tissues and cells. *Biophysical Reviews* **1**: 95–103.

[6] Wartewig S (2003). Fourier transform technique. In *IR and Raman Spectroscopy*, pages 35–51, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG.

[7] Stancik AL and Brauns EB (2008). A simple asymmetric lineshape for fitting infrared absorption spectra. *Vibrational Spectroscopy* **47**: 66–69.

[8] Bradley MS and Krech JH (1992). High-pressure Raman spectra of the acetone C-C stretch in binary liquid mixtures with methanol. *Journal of Physical Chemistry* **96**: 75–79.

[9] Kazarian SG and Chan KLA (2006). Applications of ATR-FTIR spectroscopic imaging to biomedical samples. *Biochimica et Biophysica Acta* **1758**: 858–867.

[10] Kazarian SG and Chan KLA (2010). Micro- and macro attenuated total reflection Fourier transform infrared spectroscopic imaging. *Applied Spectroscopy* **64**: 135A–152A.

[11] Chernev B, Belegratis MR and Ingolič E (2008). Investigations on multilayer films: electron microscopy and infrared spectroscopy – possibilities and limitations. *Macromolecular Symposia* **265**: 272–280.

[12] Nagle DJ, George GA, Rintoul L and Fredericks PM (2010). Use of micro-ATR/FTIR imaging to study heterogeneous polymer oxidation by direct solvent casting onto the ATR IRE. *Vibrational Spectroscopy* **53**: 24–27.

[13] Spring M, Ricci C, Peggie DA and Kazarian SG (2008). ATR-FTIR imaging for the analysis of organic materials in paint cross sections: case studies on paint samples from the National Gallery, London. *Analytical and Bioanalytical Chemistry* **392**: 37–45.

[14] Barron C, Parker ML, Mills ENC, Rouau C and Wilson RH (2005). FTIR imaging of wheat endorsperm cell walls in situ reveals compositional and architectural heterogeneity related to grain hardness. *Planta* **220**: 667–677.

[15] Holman HYN, Miles R, Hao Z, Wozei E, Anderson LM and Yang H (2009). Real-Time Chemical Imaging of Bacterial Activity in Biofilms Using Open-Channel Microfluidics and Synchrotron FTIR Spectromicroscopy. *Analytical Chemistry* **81**: 8564–8570.

[16] Miller LM, Bourassa MW and Smith RJ (2013). FTIR spectroscopic imaging of protein aggregation in living cells. *Biochimica et Biophysica Acta* **1828**: 2339–2346.

[17] Kazarian SG and Ewing AV (2013). Applications of Fourier transform infrared spectroscopic imaging to tablet dissolution and drug release. *Expert Opinion on Drug Delivery* **10**: 1207–1221.

[18] Mark S, Sahu RK, Kantarovich K, Podshyvalov A, Guterman H, Goldstein J, Jagannathan R, Argov S and Mordechai S (2004). Fourier transform infrared microspectroscopy as a quantitative diagnostic tool for assignment of premalignancy grading in cervical neoplasia. *Journal of Biomedical Optics* **9**: 558–567.

[19] Bird B, Miljković M, Remiszewski S, Akalin A, Kon M and Diem M (2012). Infrared spectral histopathology (SHP): a novel diagnostic tool for the accurate classification of lung cancer. *Laboratory Investigation* **92**: 1358–1373.

[20] Yano K, Ohoshima S, Gotou Y, Kumaido K, Moriguchi T and Katayama H (2000). Direct measurement of human lung cancerous and noncancerous tissues by

Fourier transform infrared microscopy: can an infrared microscope be used as a clinical tool? *Analytical Biochemistry* **287**: 218–225.

[21] Colagar AH, Chaichi MJ and Khadjvand T (2011). Fourier transform infrared microspectroscopy as a diagnostic tool for distinguishing between normal and malignant human gastric tissue. *Journal of Biosciences* **36**: 669–677.

[22] Gazi E, Dwyer J, Lockyer NP, Gardner P, Shanks JH, Roulson J, Hart CA, Clarke NW and Brown MD (2007). Biomolecular profiling of metastatic prostate cancer cells in bone marrow tissue using FTIR microspectroscopy: a pilot study. *Analytical and Bioanalytical Chemistry* **387**: 1621–1631.

[23] Wood BR, Quinn MA, Tait B, Ashdown M, Hislop T, Romeo M and McNaughton D (1998). FTIR microspectroscopic study of cell types and potential confounding variables in screening for cervical malignancies. *Biospectroscopy* **4**: 75–91.

[24] Sulé-Suso J, Skingsley D, Sockalingum GD, Kohler A, Kegelaer G, Manfait M and El Haj AJ (2005). FT-IR microspectroscopy as a tool to assess lung cancer cells response to chemotherapy. *Vibrational Spectroscopy* **38**: 179–184.

[25] Patel II, Harrison WJ, Kerns JG, Filik J, Wehbe K, Carmichael PL, Scott AD, Philpott MP, Frogley MD, Cinque G and Martin FL (2012). Isolating stem cells in the interfollicular epidermis employing synchrotron radiation-based Fourier-transform infrared microspectroscopy and focal plane array imaging. *Analytical and Bioanalytical Chemistry* **404**: 1745–1758.

[26] Zhao R, Quaroni L and Casson AG (2010). Fourier transform infrared (FTIR) spectromicroscopic characterization of stem-like cell populations in human esophageal normal and adenocarcinoma cell lines. *Analyst* **135**: 53–61.

[27] Gazi E, Dwyer J, Lockyer NP, Miyan J, Gardner P, Hart CA, Brown MD and Clarke NW (2005). A study of cytokinetic and motile prostate cancer cells using synchrotron-based FTIR microspectroscopic imaging. *Vibrational Spectroscopy* **38**: 193–201.

[28] Krafft C and Sergo V (2006). Biomedical applications of Raman and infrared spectroscopy to diagnose tissues. *Spectroscopy* **20**: 195–218.

[29] Ellis DI, Dunn WB, Griffin JL, Allwood JW and Goodacre R (2007). Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics* **8**: 1243–1266.

[30] Lloyd GR, Orr LE, Christie-Brown J, McCarthy K, Rose S, Thomas M and Stone N (2013). Discrimination between benign, primary and secondary malignancies in lymph nodes from the head and neck utilising Raman spectroscopy and multivariate analysis. *Analyst* **138**: 3900–3908.

[31] Gajjar K, Heppenstall LD, Pang W, Ashton KM, Trevisan J, Patel II, Llabjani V, Stringfellow HF, Martin-Hirsch PL, Dawson T and Martin FL (2013). Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis. *Analytical Methods* **5**: 89–102.

[32] Krishna CM, Kegelaer G, ADT I, Rubin S, Kartha VB, Manfait M and Sockalingum GD (2006). Combined Fourier transform infrared and raman spectroscopic approach for identification of multidrug resistance phenotype in cancer cell lines. *Biopolymers* **82**: 462–470.

[33] Crow P, Barrass B, Kendall C, Hart-Prieto M, Wright M, Persad R and Stone N (2005). The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines. *British Journal of Cancer* **92**: 2166–2170.

[34] Krafft C, Codrich D, Pelizzo G and Sergo V (2008). Raman and FTIR imaging of lung tissue: methodology for control samples. *Vibrational Spectroscopy* **46**: 141–149.

[35] Krishna CM, Sockalingum GD, Bhat RA, Venteo L, Kushtagi P, Pluot M and Manfait M (2007). FTIR and Raman microspectroscopy of normal, benign, and malignant formalin-fixed ovarian tissues. *Analytical and Bioanalytical Chemistry* **387**: 1649–1656.

[36] Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR and Feld MS (2005). Diagnosing breast cancer by using Raman spectroscopy. *Proceedings of the National Academy of Science of the United States of America* **102**: 12371–12376.

[37] Huang Z, McWilliams A, Lui H, McLean DI, Lam S and Zeng H (2003). Near-infrared Raman spectroscopy for the optical diagnosis of lung cancer. *International Journal of Cancer* **107**: 1047–1052.

[38] Kendall C, Stone N, Shepherd N, Geboes K, Warren B, Bennett R and Barr H (2003). Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus. *Journal of Pathology* **200**: 602–609.

[39] Mahadevan-Jansen A, Mitchell MF, Ramanujam N, Malpica A, Thomsen S, Utzinger U and Richards-Kortum R (1998). Near-infrared Raman spectroscopy for *in vitro* detection of cervical precancers. *Photochemistry and Photobiology* **68**: 123–132.

[40] Stone N, Kendall C, Shepherd N, Crow P and Barr H (2002). Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. *Journal of Raman Spectroscopy* **33**: 564–573.

[41] Baker MJ, Trevisan J, Bassan P, Bhargava R, Butler HJ, Dorling KM, Fielden PR, Fogarty SW, Fullwood NJ, Heys KA, Hughes C, Lasch P, Martin-Hirsch PL, Obinaju B, Sockalingum GD, Sulé-Suso J, Strong RJ, Walsh MJ, Wood BR, Gardner P and Martin FL (2014). Using Fourier transform IR spectroscopy to analyze biological materials. *Nature Protocols* **9**: 1771–1791.

[42] Trevisan J, Angelov PP, Carmichael PL, Scott AD and Martin FL (2012). Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst* **137**: 3202–3215.

[43] Ami D, Mereghetti P and Maria S (2013). Multivariate analysis for Fourier transform infrared spectra of complex biological systems and processes. In de Freitas LV and de Freitas APBR (editors), *Multivariate analysis in management, engineering and the sciences*, pages 189–220, InTech, Rijeka, Croatia.

[44] Martin FL, Kelly JG, Llabjani V, Martin-Hirsch PL, Patel II, Trevisan J, Fullwood NJ and Walsh MJ (2010). Distinguishing cell types or populations based on the computational analysis of their infrared spectra. *Nature Protocols* **5**: 1748–1760.

[45] Kelly JG, Trevisan J, Scott AD, Carmichael PL, Pollock HM, Martin-Hirsch PL and Martin FL (2011). Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers. *Journal of Proteome Research* **10**: 1437–1448.

[46] Ballabio D and Consonni V (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods* **5**: 3790–9.

[47] Petrich W (2001). Mid-infrared and Raman spectroscopy for medical diagnostics. *Applied Spectroscopy Reviews* **36**: 181–237.

[48] Budínová G, Salva J and Volka K (1997). Application of molecular spectroscopy in the mid-infrared region to the determination of glucose and cholesterol in whole blood and in blood serum. *Applied Spectroscopy* **51**: 631–635.

[49] Hoşafçi G, Klein O, Oremek G and Mäntele W (2007). Clinical chemistry without reagents? An infrared spectroscopic technique for determination of clinically relevant constituents of body fluids. *Analytical and Bioanalytical Chemistry* **387**: 1815–1822.

[50] Petibois C, Gionnet K, Gonçalves M, Perromat A, Moenner M and Déléris G (2006). Analytical performances of FT-IR spectrometry and imaging for concentration measurements within biological fluids, cells, and tissues. *Analyst* **131**: 640–647.

[51] Ellis DI and Goodacre R (2006). Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *Analyst* **131**: 875–885.

[52] Channa NA, Ghangro AB, Soomro AM and Noorani L (2007). Analysis of kidney stones by FTIR spectroscopy. *JLUMHS* **6**: 66–73.

[53] Paluszkiewicz C, Jan Ściesiński and Galka M (1988). Analysis of renal stones by FTIR spectroscopy. *Mikrochimica Acta* **I**: 45–48.

[54] Sahu RK and Mordechai S (2005). Fourier transform infrared spectroscopy in cancer detection. *Future Oncology* **1**: 635–647.

[55] Mordechai S, Sahu RK, Hammody Z, Mark S, Kantarovich K, Guterman H, Podshyvalov A, Goldstein J and Argov S (2004). Possible common biomarkers from FTIR microspectroscopy of cervical cancer and melanoma. *Journal of Microscopy* **215**: 86–91.

[56] Lewis PD, Lewis KE, Ghosal R, Bayliss S, Lloyd AJ, Wills J, Godfrey R, Kloer P and Mur LAJ (2010). Evaluation of FTIR spectroscopy as a diagnostic tool for lung cancer using sputum. *BMC Cancer* **10**: 640–649.

[57] Baker MJ, Gazi E, Brown MD, Shanks JH, Gardner P and Clarke NW (2008). FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer. *British Journal of Cancer* **99**: 1859–1866.

[58] Mackanos MA and Contag CH (2009). FTIR microspectroscopy for improved prostate cancer diagnosis. *Trends in Biotechnology* **27**: 661–663.

[59] Rigas B, Morgello S, Goldman IS and Wong PTT (1990). Human colorectal cancers display abnormal Fourier-transform infrared spectra. *Proceedings of the National Academy of Science of the United States of America* **87**: 8140–8144.

[60] Mackanos MA, Hargrove J, Wolters R, Du CB, Friedland S, Soetikno RM, Contag CH, Arroyo MR, Crawford JM and Wang TD (2009). Use of an endoscope-compatible probe to detect colonic dysplasia with Fourier transform infrared spectroscopy. *Journal of Biomedical Optics* **14**: 044006–1–044006–8.

[61] Argov S, Ramesh J, Salman A, Sinelnikov I, Goldstein J, Guterman H and Mordechai S (2002). Diagnostic potential of Fourier-transform infrared microspectroscopy and advanced computational methods in colon cancer patients. *Journal of Biomedical Optics* **7**: 248–254.

[62] Kallenbach-Thieltges A, Großerüschkamp F, Mosig A, Diem M, Tannapfel A and Gerwert K (2013). Immunohistochemistry, histopathology and infrared spectral histopathology of colon cancer tissue sections. *Journal of Biophotonics* **6**: 88–100.

[63] Wang JS, Shi JS, Xu YZ, Duan XY, Zhang L, Wang J, Yang LM, Weng SF and Wu JG (2003). FT-IR spectroscopic analysis of normal and cancerous tissues of esophagus. *World Journal of Gastroenterology* **9**: 1897–1899.

[64] Maziak DE, Do MT, Shamji FM, Sundaresan SR, Perkins DG and Wong PTT (2007). Fourier-transform infrared spectroscopic study of characteristic molecular structure in cancer cells of esophagus: an exploratory study. *Cancer Detection and Prevention* **31**: 244–253.

[65] Quaroni L and Casson AG (2009). Characterization of Barrett esophagus and esophageal adenocarcinoma by Fourier-transform infrared microscopy. *Analyst* **134**: 1240–1246.

[66] Wang TD, Triadafilopoulos G, Crawford JM, Dixon LR, Bhandari T, Sahbaie P, Friedland S, Soetikno R and Contag CH (2007). Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy. *Proceedings of the National Academy of Science of the United States of America* **104**: 15864–15869.

[67] Petibois C and Déléris G (2006). Chemical mapping of tumor progression by FT-IR imaging: towards molecular histopathology. *Trends in Biotechnology* **24**: 455–462.

[68] Walsh MJ, Fellous TG, Hammiche A, Lin WR, Fullwood NJ, Grude O, Bahrami F, Nicholson JM, Cotte M, Susini J, Pollock HM, Brittan M, Martin-Hirsch PL, Alison MR and Martin FL (2008). Fourier transform infrared microspectroscopy identifies symmetric PO_2- modifications as a marker of the putative stem cell region of human intestinal crypts. *Stem Cells* **26**: 108–118.

[69] Boydston-White S, Romeo M, Chernenko T, Regina A, Miljković M and Diem M (2006). Cell-cycle-dependent variations in FTIR micro-spectra of single proliferating HeLa cells: principal component and artificial neural network analysis. *Biochimica et Biophysica Acta* **1758**: 908–914.

[70] Jessen TE, Höskuldsson AT, Bjerrum PJ, Verder H, Sørensen L, Bratholm PS, Christensen B, Jensen LS and Jensen MAB (2014). Simultaneous determination of glucose, triglycerides, urea, cholesterol, albumin and total protein in human plasma by Fourier transform infrared spectroscopy: direct clinical biochemistry without reagents. *Clinical Biochemistry* **47**: 1306–1312.

[71] Low-Ying S, Shaw RA, Leroux M and Mantsch HH (2002). Quantitation of glucose and urea in whole blood by mid-infrared spectroscopy of dry films. *Vibrational Spectroscopy* **28**: 111–116.

[72] Heise HM and Bittner A (1995). Multivariate calibration for physiological samples using infrared spectra with choice of different intensity data. *Journal of Molecular Structure* **348**: 127–130.

[73] Petrich W, Dolenko B, Früh J, Ganz M, Greger H, Jacob S, Keller F, Nikulin AE, Otto M, Quarder O, Somorjai RL, Staib A, Werner G and Wielinger H (2000). Disease pattern recognition in infrared spectra of human sera with diabetes mellitus as an example. *Applied Optics* **39**: 3372–3379.

[74] Lacombe C, Untereiner V, Gobinet C, Zater M, Sockalingum GD and Garnotel R (2015). Rapid screening of classic galactosemia patients: a proof-of-concept study using high-throughput FTIR analysis of plasma. *Analyst* **140**: 2280–2286.

[75] Scaglia E, Sockalingum GD, Schmitt J, Gobinet C, Schneider N, Manfait M and Thiéfin G (2011). Noninvasive assessment of hepatic fibrosis in patients with chronic hepatitis C using serum Fourier transform infrared spectroscopy. *Analytical and Bioanalytical Chemistry* **401**: 2919–2925.

[76] Fabian H, Lasch P and Naumann D (2005). Analysis of biofluids in aqueous environment based on mid-infrared spectroscopy. *Journal of Biomedical Optics* **10**: 031103–1–031103–10.

[77] Ollesch J, Drees SL, Heise HM, Behrens T, Brüning T and Gerwert K (2013). FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification. *Analyst* **138**: 4092–4102.

[78] Hands JR, Abel P, Ashton K, Dawson T, Davis C, Lea RW, McIntosh AJS and Baker MJ (2013). Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors. *Analytical and Bioanalytical Chemistry* **405**: 7347–7355.

[79] Gajjar K, Trevisan J, Owens G, Keating PJ, Wood NJ, Stringfellow HF, Martin-Hirsch PL and Martin FL (2013). Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer. *Analyst* **138**: 3917–10.

[80] Khanmohammadi M, Garmarudi AB, Ramin M and Ghasemi K (2013). Diagnosis of renal failure by infrared spectrometric analysis of human serum samples and soft independent modeling of class analogy. *Microchemical Journal* **106**: 67–72.

[81] Heise HM, Voigt G, Lampen P, Küpper L, Rudloff S and Werner G (2001). Multivariate calibration for the determination of analytes in urine using mid-infrared attenuated total reflection spectroscopy. *Applied Spectroscopy* **55**: 434–443.

[82] Markus APJA, Swinkels DW, Jakobs BS, Wevers RA, Trijbels JMF and Willems HL (2001). New technique for diagnosis and monitoring of alcaptonuria: quantification of homogentisic acid in urine with mid-infrared spectrometry. *Analytica Chimica Acta* **429**: 287–292.

[83] Shaw RA, Rigatto C, Reslerova M, Ying SL, Man A, Schattka B, Battrell CF, Matthewson J and Mansfield C (2009). Toward point-of-care diagnostic metabolic fingerprinting: quantification of plasma creatinine by infrared spectroscopy of microfluidic-preprocessed samples. *Analyst* **134**: 1224–1231.

[84] Mulready KJ and McGoldrick D (2012). The establishment of a standard and real patient kidney stone library utilizing Fourier transform-infrared spectroscopy with a diamond ATR accessory. *Urological Research* **40**: 483–498.

[85] Baker MJ, Gazi E, Brown MD, Shanks JH, Clarke NW and Gardner P (2009). Investigating FTIR based histopathology for the diagnosis of prostate cancer. *Journal of Biophotonics* **2**: 104–113.

[86] Bird B, Romeo M, Laver N and Diem M (2009). Spectral detection of micrometastases in lymph node histo-pathology. *Journal of Biophotonics* **2**: 37–46.

[87] Yano K, Ohoshima S, Shimizu Y, Moriguchi T and Katayama H (1996). Evaluation of glycogen level in human lung carcinoma tissues by an infrared spectroscopic method. *Cancer Letters* **110**: 29–34.

[88] Lee SY, Yoon KA, Jang SH, Ganbold EO, Uuriintuya D, Shin SM, Ryu PD and Joo SW (2009). Infrared spectroscopy characterization of normal and lung cancer cells originated from epithelium. *Journal of Veterinary Science* **10**: 299–304.

[89] Sun X, Xu Y, Wu J, Zhang Y and Sun K (2013). Detection of lung cancer tissue by attenuated total reflection-Fourier transform infrared spectroscopy-a pilot study of 60 samples. *Journal of Surgical Research* **179**: 33–38.

[90] Walsh MJ, Kajdacsy-Balla A, Holton SE and Bhargava R (2012). Attenuated total reflectance Fourier-transform infrared spectroscopic imaging for breast histopathology. *Vibrational Spectroscopy* **60**: 23–28.

[91] Shaw RA, Low-Ying S, Leroux M and Mantsch HH (2000). Toward reagent-free clinical analysis: quantitation of urine urea, creatinine, and total protein from the mid-infrared spectra of dried urine films. *Clinical Chemistry* **46**: 1493–1495.

[92] Greenspan L (1977). Humidity fixed points of binary saturated aqueous solutions. *Journal of Research of the National Bureau of Standards - A Physics and Chemistry* **81A**: 89–96.

[93] Goormaghtigh E and Ruysschaert JM (1994). Subtraction of atmospheric water contribution in Fourier transform infrared spectroscopy of biological membranes and proteins. *Spectrochimica Acta* **50A**: 2137–2144.

[94] Chalmers JM (2006). *Handbook of vibrational spectroscopy*. Mid-infrared spectroscopy: anomalies and common errors, John Wiley & Sons, Ltd., Chichester, UK.

[95] Rinnan Å, Berg Fvd and Engelsen SB (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry* **28**: 1201–1222.

[96] Lasch P (2012). Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemometrics and Intelligent Laboratory Systems* **117**: 100–114.

[97] Knapp EW and Fischer SF (1982). The concentration dependence of the vibrational linewidth and shift in liquid binary mixtures: An analytical model. *Journal of Chemical Physics* **76**: 4730–4735.

[98] Sánchez-Bajo F and Cumbrera FL (1997). The use of the pseudo-Voigt function in the variance method of x-ray line-broadening analysis. *Journal of Applied Crystallography* **30**: 427–430.

[99] Bland JM and Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**: 307–310.

[100] Suenaga M (2005). Facio: New Computational Chemistry Environment for PC GAMESS. *Journal of Computer Chemistry Japan* **4**: 25–32.

[101] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X and Hratchian HP (2009). *Gaussian 09, revision A. 1.* Wallingford, CT, USA.

[102] Tantirungrotechai Y, Phanasant K, Roddecha S, Surawatanawong P, Sutthikhum V and Limtrakul J (2006). Scaling factors for vibrational frequencies and zero-point vibrational energies of some recently developed exchange-correlation functionals. *Journal of Molecular Structure* **760**: 189–192.

[103] Moore S, Spackman DH and Stein WH (1958). Chromatography of amino acids on sulfonated polystyrene resins. *Analytical Chemistry* **30**: 1185–1190.

[104] Guerra A, Petrarulo M, Schianchi T, Allegri F, Meschi T, Bruno M, Ramello A, Baruffaldi M, Novarini A and Borghi L (2002). A simple quantitative test for screening cystinuria. *Laboratory Medicine* **33**: 214–217.

[105] Putnam DF (1971). Composition and concentrative properties of human urine. National Aeronautics and Space Administration, USA.

[106] Rose C, Parker A, Jefferson B and Cartmell E (2015). The characterization of feces and urine: a review of the literature to inform advanced treatment technology. *Critical Reviews in Environmental Science and Technology* **45**: 1827–1879.

[107] Barr DB, Wilder LC, Caudill SP, Gonzalez AJ, Needham LL and Pirkle JL (2004). Urinary creatinine concentrations in the U.S. population: implications for urinary biological monitoring measurements. *Environmental Health Perspectives* **113**: 192–200.

[108] Simerville JA, Maxted WC and Pahira JJ (2005). Urinalysis: a comprehensive review. *American Family Physician* **71**: 1153–1162.

[109] Keuleers R, Desseyn O, Rousseau B and Van Alsenoy C (1999). Vibrational analysis of urea. *Journal of Physical Chemistry A* **103**: 4621–4630.

[110] Rousseau B, Van Alsenoy C, Keuleers R and Desseyn HO (1998). Solids modeled by ab-initio crystal field methods. Part 17. Study of the structure and vibrational spectrum of urea in the gas phase and in its *P42 1m* crystal phase. *Journal of Physical Chemistry A* **102**: 6540–6548.

[111] Knepper MA, Packer R and Good DW (1989). Ammonium transport in the kidney. *Physiological reviews* **69**: 179–249.

[112] Isabelle M, Stone N, Barr H, Vipond M, Shepherd N and Rogers K (2008). Lymph node pathology using optical spectroscopy in cancer diagnostics. *Spectroscopy* **22**: 97–104.

[113] Wolthuis R, Travo A, Nicolet C, Neuville A, Gaub MP, Guenot D, Ly E, Manfait M, Jeannesson P and Piot O (2008). IR spectral imaging for histopathological characterization of xenografted human colon carcinomas. *Analytical Chemistry* **80**: 8461–8469.

[114] Diem M (2015). *Modern vibrational spectroscopy & micro-spectroscopy*. Theory, instrumentation & biomedical applications, John Wiley & Sons Ltd., Northeastern University, USA.

[115] Moss D (2010). *Biomedical applications of synchrotron infrared microspectroscopy*. A practical approach, Royal Society of Chemistry, Cambridge, UK.

[116] Petibois C, Rigalleau V, Melin AM, Perromat A, Cazorla G, Gin H and Déléris G (1999). Determination of glucose in dried serum samples by Fourier-transform Infrared spectroscopy. *Clinical Chemistry* **45**: 1530–1535.

[117] Pérez-Guaita D, Sánchez-Illana Á, Garrigues S and de la Guardia M (2015). Determination of lidocaine in urine at low ppm levels using dispersive microextraction and attenuated total reflectance–Fourier transform infrared measurements of dry films. *Microchemical Journal* **121**: 178–183.

[118] Stumpe MC and Grubmüller H (2007). Aqueous urea solutions: structure, energetics and urea aggregation. *Journal of Physical Chemistry B* **111**: 6220–6228.

[119] Lee C, Stahlberg EA and Fitzgerald G (1995). Chemical structure of urea in water. *Journal of Physical Chemistry* **99**: 17737–17741.

[120] Grdadolnik J and Maréchal Y (2002). Urea and urea–water solutions—an infrared study. *Journal of Molecular Structure* **615**: 177–189.

[121] Nandel FS, Verma R, Singh B and Jain DVS (1998). Mechanism of hydration of urea and guanidium ion: a model study of denaturation of proteins. *Pure and Applied Chemistry* **70**: 659–664.

[122] Rezus YLA and Bakker HJ (2006). Effect of urea on the structural dynamics of water. *Proceedings of the National Academy of Science of the United States of America* **103**: 18417–18420.

[123] Carr J, Buchanan LE, Schmidt JR, Zanni MT and Skinner JL (2013). Structure and dynamics of urea/water mixtures investigated by vibrational spectroscopy and molecular dynamics simulation. *Journal of Physical Chemistry B* **117**: 13291–13300.

[124] Finer EG, Franks F and Tait MJ (1972). Nuclear magnetic resonance studies of aqueous urea solutions. *Journal of the American Chemical Society* **94**: 4424–4429.

[125] Jung YM, Czarnik-Matusewicz B and Bin Kim S (2004). Characterization of concentration dependent IR spectral variations of urea aqueous solutions by PCA and 2D correlation spectroscopy. *Journal of Physical Chemistry B* **108**: 13008–13014.

[126] Tanaka H, Nakanishi K and Touhara H (1985). Computer experiments on aqueous-solutions .7. Potential energy function for urea dimer and molecular dynamics

calculation of 8 mol-percent aqueous solution of urea. *Journal of Chemical Physics* **82**: 5184–5191.

[127] Hoccart X and Turrell G (1993). Raman spectroscopic investigation of the dynamics of urea–water complexes. *Journal of Chemical Physics* **99**: 8498–8503.

[128] Idrissi A (2005). Molecular structure and dynamics of liquids: aqueous urea solutions. *Spectrochimica Acta Part A* **61**: 1–17.

[129] Trendafilova N, Kurbakova AP, Efimenko IA, Mitewa M and Bontchev PR (1991). Infrared spectra of Pt(II) creatinine complexes. Normal coordinate analysis of creatinine and Pt(creat)2(NO2)2. *Spectrochimica Acta* **47**: 577–584.

[130] Bayrak C and Bayari SH (2010). Vibrational and DFT studies of creatinine and its metal complexes. *Hacettepe Journal of Biology and Chemistry* **38**: 107–118.

[131] Magee EA, Curno R, Edmond LM and Cummings JH (2004). Contribution of dietary protein and inorganic sulfur to urinary sulfate: toward a biomarker of inorganic sulfur intake. *American Journal of Clinical Nutrition* **80**: 137–142.

[132] Takeda E, Taketani Y, Sawada N, Sato T and Yamamoto H (2004). The regulation and function of phosphate in the human body. *BioFactors* **21**: 345–355.

[133] Biyani CS and Cartledge JJ (2006). Cystinuria—diagnosis and management. *EAU-EBU Update Series* **4**: 175–183.

[134] Furlan P, Servey J, Scott S and Peaslee M (2004). FTIR analysis of mouse urine urea using IR cards. *Spectroscopy Letters* **37**: 311–318.

[135] Chillarón J, Font-Llitjós M, Fort J, Zorzano A, Goldfarb DS, Nunes V and Palacín M (2010). Pathophysiology and treatment of cystinuria. *Scientific Reports* **6**: 424–434.

[136] Carta R and Tola G (1996). Solubilities of L-cystine, L-tyrosine, L-leucine, and glycine in aqueous solutions at various pHs and NaCl concentrations. *Journal of Chemical Engineering Data* **41**: 414–417.

[137] Dent CE and Senior B (1955). Studies on the treatment of cystinuria. *British Journal of Urology* **27**: 317–332.

[138] Goodyer P, Saadi I, Ong P, Elkas G and Rozen R (1998). Cystinuria subtype and the risk of nephrolithiasis. *Kidney International* **54**: 56–61.

[139] Finocchiaro R, D'Eufemia P, Celli M, Zaccagnini M, Viozzi L, Troiani P, Mannarino O and Giardini O (1998). Usefulness of cyanide-nitroprusside test in detecting incomplete recessive heterozygotes for cystinuria: a standardized dilution procedure. *Urological Research* **26**: 401–405.

[140] Shinohara K and Padis KE (1936). The determination of thiol and disulphide compounds, with special reference to cysteine and cystine. *Journal of Biological Chemistry* **112**: 709–721.

[141] Goldfarb DS, Coe FL and Asplin JR (2006). Urinary cystine excretion and capacity in patients with cystinuria. *Kidney International* **69**: 1041–1047.

[142] Marickar YMF, Lekshmi PR, Varma L and Koshy P (2009). Problem in analysing cystine stones using FTIR spectroscopy. *Urological Research* **37**: 263–269.

[143] Vasiliades J (1976). Reaction of alkaline sodium picrate with creatinine: I. Kinetics and mechanism of formation of the mono-creatinine picric acid complex. *Clinical Chemistry* **22**: 1664–1671.

[144] Bonsnes RW and Taussky HH (1958). On the colorimetric determination of creatinine by the Jaffe reaction. *Journal of Biological Chemistry* **158**: 581–591.

[145] Edvardsson V, Palsson R, Olafsson I, Hjaltadottir G and Laxdal T (2001). Clinical features and genotype of adenine phosphoribosyltransferase deficiency in Iceland. *American Journal of Kidney Diseases* **38**: 473–480.

[146] Cassidy MJD, McCulloch T, Fairbanks LD and Simmonds HA (2004). Diagnosis of adenine phosphoribosyltransferase deficiency as the underlying cause of renal failure in a renal transplant recipient. *Nephrology Dialysis Transplantation* **19**: 736–738.

[147] Peck CC, Bailey FJ and Moore GL (1977). Enhanced solubility of 2,8 dihydroxyadenine (DOA) in human urine. *Transfusion* **17**: 383–390.

[148] Mitchell SC and Smith RL (2001). Trimethylaminuria: the fish malodor syndrome. *Drug Metabolism and Disposition* **29**: 517–521.

[149] Wise PM, Eades J, Tjoa S, Fennessey PV and Preti G (2011). Individuals reporting idiopathic malodor production: demographics and incidence of trimethylaminuria. *American Journal of Medicine* **124**: 1058–1063.

[150] Lindbäck B and Bergman A (1989). A new commercial method for the enzymatic determination of creatinine in serum and urine evaluated: comparison with a kinetic Jaffé method and isotope dilution-mass spectrometry. *Clinical Chemistry* **35**: 835–837.

[151] Weber JA and van Zanten AP (1991). Interferences in current methods for measurements of creatinine. *Clinical Chemistry* **37**: 695–700.

[152] Ngo TC and Assimos DG (2007). Uric acid nephrolithiasis: recent progress and future directions. *Reviews in urology* **9**: 17–27.

[153] Dorling KM and Baker MJ (2013). Highlighting attenuated total reflection Fourier transform infrared spectroscopy for rapid serum analysis. *Trends in Biotechnology* **31**: 327–328.

[154] Khanmohammadi M, Nasiri R, Ghasemi K, Samani S and Bagheri Garmarudi A (2007). Diagnosis of basal cell carcinoma by infrared spectroscopy of whole blood samples applying soft independent modeling class analogy. *Journal of Cancer Research and Clinical Oncology* **133**: 1001–1010.

[155] Mitchell AL, Gajjar KB, Theophilou G, Martin FL and Martin-Hirsch PL (2014). Vibrational spectroscopy of biofluids for disease screening or diagnosis: translation from the laboratory to a clinical setting. *Journal of Biophotonics* **7**: 153–165.

[156] Korenkova V, Jones A, Hoy WE, Morais C, Cooper MA and Gobe GC (2015). Urinary biomarkers for detection of early and advanced chronic kidney disease - a pilot study. *Medicinal Chemistry* **5**: 96–103.

[157] Vaidya VS, Waikar SS, Ferguson MA, Collings FB, Sunderland K, Gioules C, Bradwin G, Matsouaka R, Betensky RA, Curhan GC and Bonventre JV (2008). Urinary biomarkers for sensitive and specific detection of acute kidney injury in humans. *Clinical and Translational Science* **1**: 200–208.

[158] Koyner JL, Vaidya VS, Bennett MR, Ma Q, Worcester E, Akhter SA, Raman J, Jeevanandam V, O'Connor MF, Devarajan P, Bonventre JV and Murray PT (2010). Urinary biomarkers in the clinical prognosis and early detection of acute kidney injury. *Clinical Journal of the American Society of Nephrology : CJASN* **5**: 2154–2165.

[159] Parker J and Spiess PE (2011). Current and emerging bladder cancer urinary biomarkers. *Scientific World Journal* **11**: 1103–1112.

[160] Lam T and Nabi G (2007). Potential of urinary biomarkers in early bladder cancer diagnosis. *Expert Review of Anticancer Therapy* **7**: 1105–1115.

[161] Zhang H, Cao J, Li L, Liu Y, Zhao H, Li N, Li B, Zhang A, Huang H, Chen S, Dong M, Yu L, Zhang J and Chen L (2015). Identification of urine protein biomarkers with the potential for early detection of lung cancer. *Scientific Reports* **5**: 11805.

[162] Bauer C, Melamed ML and Hostetter TH (2008). Staging of chronic kidney disease: time for a course correction. *Journal of the American Society of Nephrology* **19**: 844–846.

[163] Hemmelgarn BR, Manns BJ, Lloyd A, James MT, Klarenbach S, Quinn RR, Wiebe N, Tonelli M and Alberta Kidney Disease Network (2010). Relation between kidney function, proteinuria, and adverse outcomes. *Journal of the American Medical Association* **303**: 423–429.

[164] Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, Hogg RJ, Perrone RD, Lau J, Eknoyan G and National Kidney Foundation (2003). National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Annals of Internal Medicine* **139**: 137–147.

[165] Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J and CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) (2009). A new equation to estimate glomerular filtration rate. *Annals of Internal Medicine* **150**: 604–612.

[166] McClellan WM, Knight DF, Karp H and Brown WW (1997). Early detection and treatment of renal disease in hospitalized diabetic and hypertensive patients: important differences between practice and published guidelines. *American Journal of Kidney Diseases* **29**: 368–375.

[167] Obrador GT, Ruthazer R, Arora P, Kausz AT and Pereira BJ (1999). Prevalence of and factors associated with suboptimal care before initiation of dialysis in the United States. *Journal of the American Society of Nephrology* **10**: 1793–1800.

[168] Hanahan D and Weinberg RA (2011). Hallmarks of cancer: the next generation. *Cell* **144**: 646–674.

[169] Weinberg R (2013). *The biology of cancer*. Garland Science, Taylor and Francis Group, LLC, New York, USA, 2nd edition.

[170] Cascinu S, Falconi M, Valentini V, Jelic S and ESMO Guidelines Working Group (2010). Pancreatic cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* **21**: 55–58.

[171] Siegel R, Naishadham D and Jemal A (2013). Cancer statistics, 2013. *CA-A Cancer Journal for Clinicians* **63**: 11–30.

[172] Fesinmeyer MD, Austin MA, Li CI, De Roos AJ and Bowen DJ (2005). Differences in survival by histologic type of pancreatic cancer. *Cancer Epidemiology, Biomarkers & Prevention* **14**: 1766–1773.

[173] Bellizzi AM and Frankel WL (2009). Pancreatic pathology: a practical review. *Laboratory Medicine* **40**: 417–426.

[174] Chari ST (2007). Detecting early pancreatic cancer: problems and prospects. *Seminars in Oncology* **34**: 284–294.

[175] Li D, Xie K, Wolff R and Abbruzzese JL (2004). Pancreatic cancer. *Lancet* **363**: 1049–1057.

[176] Lowenfels AB, Maisonneuve P, Cavallini G, Ammann RW, Lankisch PG, Andersen JR, Dimagno EP, Andrén-Sandberg A and Domellöf L (1993). Pancreatitis and the risk of pancreatic cancer. *New England Journal of Medicine* **328**: 1433–1437.

[177] Krejs GJ (2010). Pancreatic cancer: epidemiology and risk factors. *Digestive Diseases* **28**: 355–358.

[178] Hruban RH and Fukushima N (2007). Pancreatic adenocarcinoma: update on the surgical pathology of carcinomas of ductal origin and PanINs. *Modern Pathology* **20 Suppl 1**: S61–70.

[179] Lee ES and Lee JM (2014). Imaging diagnosis of pancreatic cancer: a state-of-the-art review. *World Journal of Gastroenterology* **20**: 7864–7877.

[180] Kondepati VR, Keese M, Heise HM and Backhaus J (2006). Detection of structural disorders in pancreatic tumour DNA with Fourier-transform infrared spectroscopy. *Vibrational Spectroscopy* **40**: 33–39.

[181] Chen YJ, Cheng YD, Liu HY, Lin PY and Wang CS (2006). Observation of biochemical imaging changes in human pancreatic cancer tissue using Fourier-transform infrared microspectroscopy. *Chang Gun Medical Journal* **29**: 518–527.

[182] Szajda SD, Waszkiewicz N, Chojnowska S and Zwierz K (2011). Carbohydrate markers of pancreatic cancer. *Biochemical Society Transactions* **39**: 340–343.

[183] Fong ZV and Winter JM (2012). Biomarkers in pancreatic cancer: diagnostic, prognostic, and predictive. *Cancer Journal* **18**: 530–538.

[184] Durlik M and Tuchalska-Czuroń J (2014). Ploidy and DNA index as prognostic factors in resected pancreatic ductal adenocarcinoma - a review of the literature. *Przeglad gastroenterologiczny* **9**: 313–316.

[185] Jenkinson C, Earl J, Ghaneh P, Halloran C, Carrato A, Greenhalf W, Neoptolemos J and Costello E (2015). Biomarkers for early diagnosis of pancreatic cancer. *Expert Review of Gastroenterology & Hepatology* **9**: 305–315.

[186] Barlogie B, Raber MN, Schumann J, Johnson TS, Drewinko B, Swartzendruber DE, Göhde W, Andreeff M and Freireich EJ (1983). Flow cytometry in clinical cancer research. *Cancer Research* **43**: 3982–3997.

[187] Jonckheere N, Skrypek N and Van Seuningen I (2010). Mucins and pancreatic cancer. *Cancers* **2**: 1794–1812.

[188] Wang G, Lipert RJ, Jain M, Kaur S, Chakraboty S, Torres MP, Batra SK, Brand RE and Porter MD (2011). Detection of the potential pancreatic cancer marker MUC4 in serum using surface-enhanced Raman scattering. *Analytical Chemistry* **83**: 2554–2561.

[189] Ohlund D, Lundin C, Ardnor B, Oman M, Naredi P and Sund M (2009). Type IV collagen is a tumour stroma-derived biomarker for pancreas cancer. *British Journal of Cancer* **101**: 91–97.

[190] Kondepati VR, Zimmermann J, Keese M, Sturm J, Manegold BC and Backhaus J (2005). Near-infrared fiber optic spectroscopy as a novel diagnostic tool for the detection of pancreatic cancer. *Journal of Biomedical Optics* **10**: 054016–1–054016–6.

Appendix A

Benchtop spectrometer instructions

Instructions for recording infrared spectra on bench top Alpha ATR-FTIR spectrometer.

The Alpha machine should be on permanently and Opus can be left running (skip to step 5 if Opus is already running). However, please switch off the monitor when not in use to extend its lifetime.

Starting Up Opus

1. Upon computer start-up open Opus 7 by double clicking the icon on the desk top
2. Enter password **OPUS** (in capitals) when prompted
3. When Opus starts up a license window appears, click 'OK'.
4. Wait for performance tests to finish (status bar at bottom of screen will be green when tests are running and grey when finished).

Check Status

5. Check that traffic light symbol in bottom right-hand corner of main Opus workspace is green.

To Record Background (once a day)

6. Clean prism with water and ethanol and wait until prism is dry.
7. Click 'REFRESH BACKGROUND'.
8. Click 'Continue' to run background. Wait until background has finished, indicated by grey status bar (approx. 2 minutes).

To Run Sample Urine Spectrum

9. Click 'RECORD SPECTRUM'.
10. Clean prism (as above) and click 'Continue'.
11. Enter sample name (and additional information if required). Click 'Continue'.
12. Before putting a sample on the prism click 'Continue'. This will bring up a live spectrum of the prism surface.
13. If you are happy that the prism is clean place 50 µL of sample urine on to prism. Ensure entire prism is covered. Click the 'Start Measurement' button at the bottom left-hand side of the spectrum window.
    a. If the spectrum is not flat click the 'Abort' button at the bottom left-hand side of the spectrum window (close the 3 error messages that pop up) and run a new background (steps 6-8).
14. Wait for sample scan to finish (approx. 1 minute).
15. To take further measurements, clean the prism with water and ethanol, and follow the protocol steps 9-14.

Instruction Manual

Bruker Alpha FTIR Spectrometer

Contents

1. **Starting up Opus software** (p. 2)
2. **Running performance checks** (p. 4)
3. **The spectrometer** (p. 7)
4. **Cleaning the prism** (p. 9)
5. **Refreshing the background** (p. 9)
6. **Recording a sample spectrum** (p. 11)
7. **Loading an Opus spectrum** (p. 15)
8. **Troubleshooting** (p. 17)
9. **Contact details** (p.17)

1. Starting up Opus

Opus can be left running all the time. In the event that you find Opus has been closed, double left click on the 'Opus 7.0' icon on the desktop.

Figure 1: Opus 7.0 shortcut on desktop

Upon opening Opus you will be prompted to input a password:

User ID: **Default**

Password : **OPUS**

Assigned workspaces: **C:\OPUS_7.2.139.1294\RFHfull_access.ows**

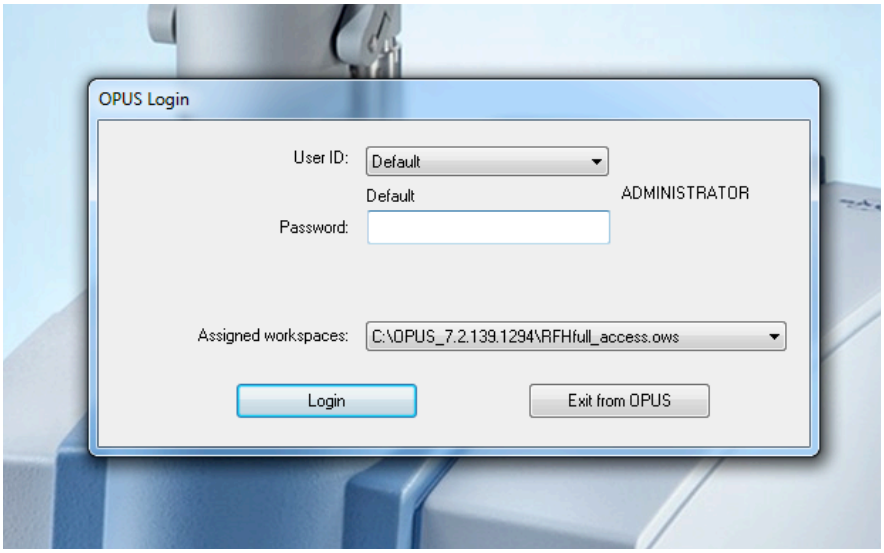


Figure 2: Opus password prompt.

On this page you will usually only have to put in the password. You will only need to change the user ID or assigned workspace if the user before you has changed these settings.

Figure 3 shows the layout of the main Opus workspace .

3

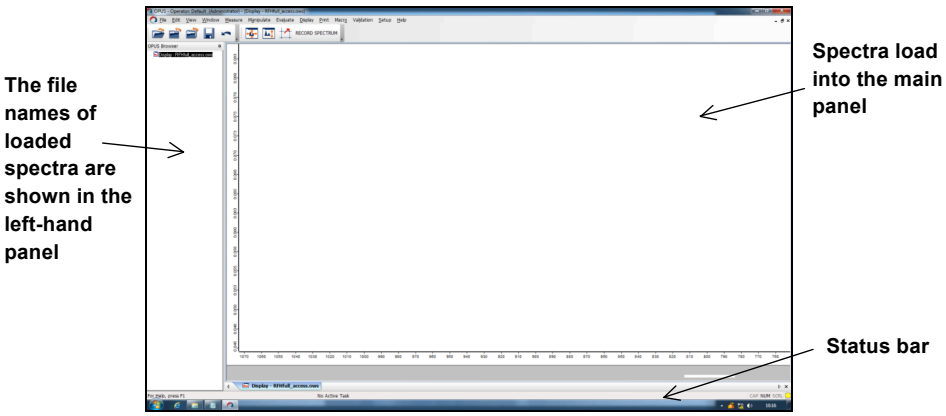


Figure 3: Main Opus workspace. Red box indicates location of status light.

Opus usually automatically runs some performance tests upon Opus start up. Please wait for these to be complete before performing any other tasks. A window will appear in the bottom right-hand corner of the screen once the tests are complet click close (Fig. 4). If the status bar reads 'No active task' then you can continue with set up.

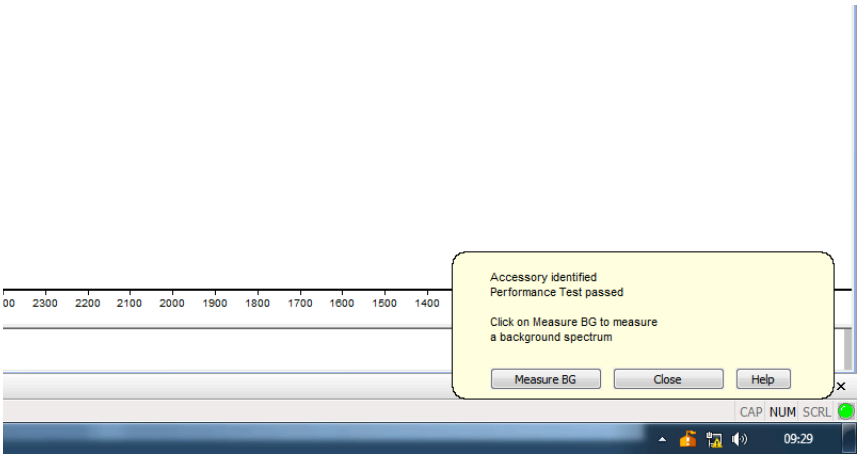


Figure 4: Window reporting on results of start-up performance tests.

2. Running performance checks

There is a traffic light symbol in the bottom right-hand corner for the main Opus workspace (Fig. 5). This should always be green during operation. If the light is yellow or red, left click on it and a diagnostics window will open (entitled “Instrument Status”, Fig. 6). Typically, the traffic light will be yellow if the diagnostics tests are out of date.

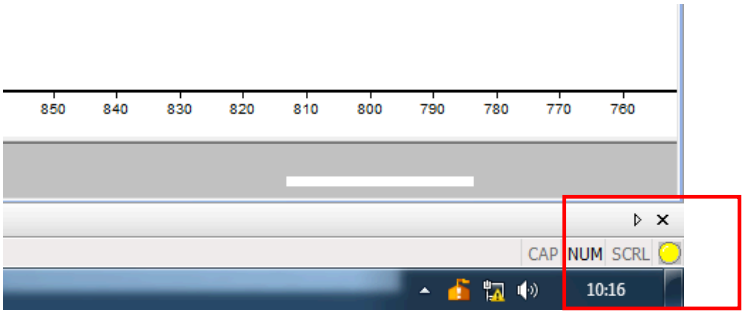


Figure 5: Traffic light symbol.

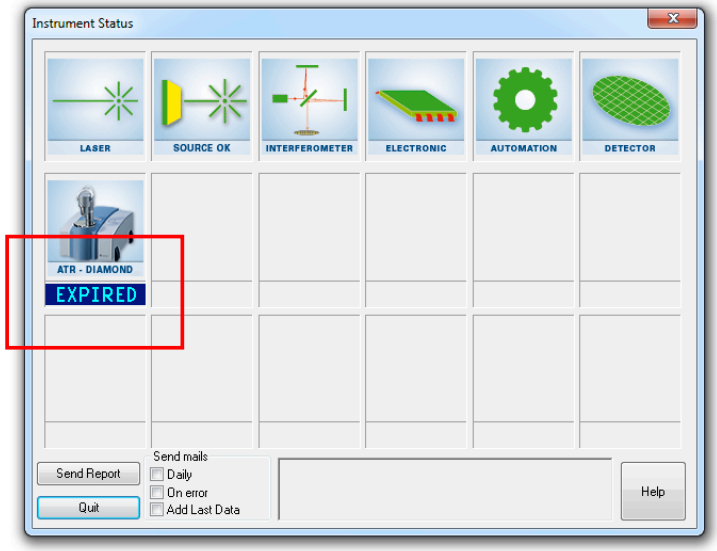


Figure 6: Diagnostics window that appears after clicking on the traffic light symbol. Red box indicates warning showing that diagnostic tests are out of date.

5

The ‘Instrument Status’ window shows several boxes corresponding to difference aspects of the machine. If there is a problem with any of these an error message will be shown underneath the symbol for that component.

Underneath the ‘ATR-Diamond’ box it will say ‘EXPIRED’ if the diagnostic tests are out of date. In this case, left click on the box and a further dialog window will open (Fig. 7).

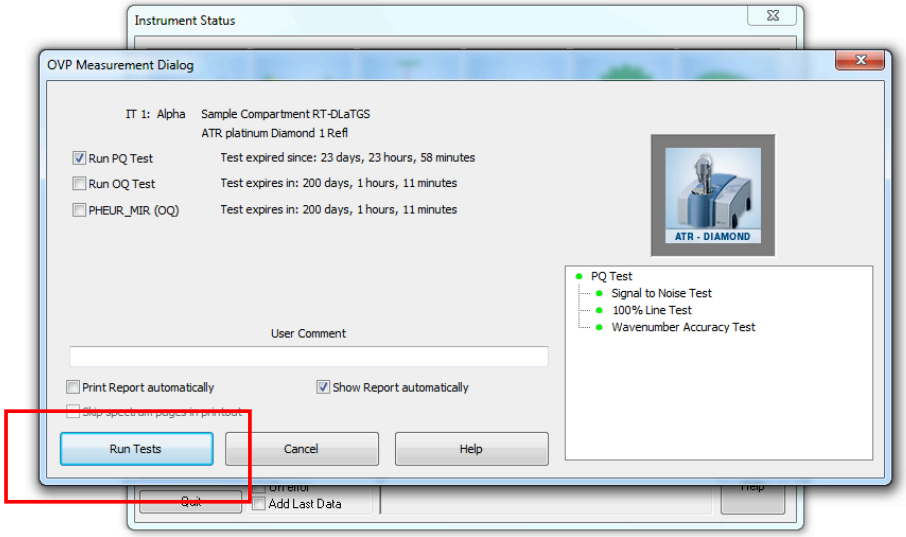


Figure 7: Run diagnostics window.

Ordinarily you do not need to change any settings on this page, simply left click on ‘Run Tests’ and the diagnostic tests will begin. Usually you will only need to run a PQ test, however, if either OQ test is reported as expired then ensure that the check boxes for these tests are also ticked before clicking ‘Run Tests’. A status bar at the bottom of the main Opus workspace will indicate the progress of these tests. Whilst the status bar is green (Fig. 8 top), tests are still running and the machine should be left alone. If the status bar is grey and reads ‘No active task’ (Fig. 8 bottom), all processes are finished and you can proceed with measurements.

6

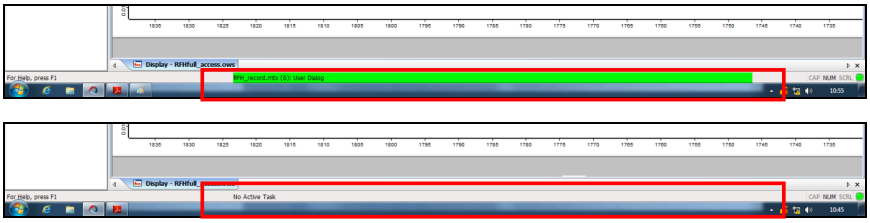


Figure 8: Status bar showing that a process is running (top) and when no tasks are running (below).

After the diagnostic tests are finished a report will load. This is a PDF file that is saved automatically. All you need to do is check that the diagnostic tests were passed and then close this window (Fig. 9).

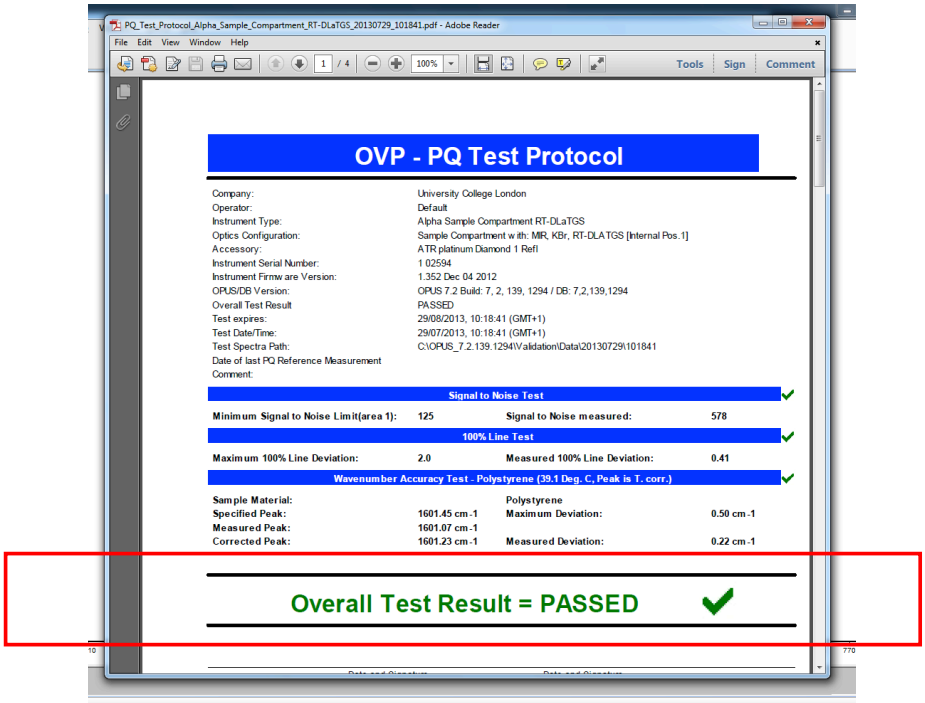


Figure 9: Diagnostic tests report. Red box highlights the overall result of the test.

3. The spectrometer

Figure 17 shows the Alpha machine and the computer. Figure 11 shows the position of the prism on the machine. This prism is the active area and where you should load your sample. It is important that the entire prism is covered with sample (the prism is only a few millimetres square so this should be no problem. The light (indicated by the red box in Fig. 10) should be green during operation. If it is not green this indicates that there is a problem with the spectrometer or the software.

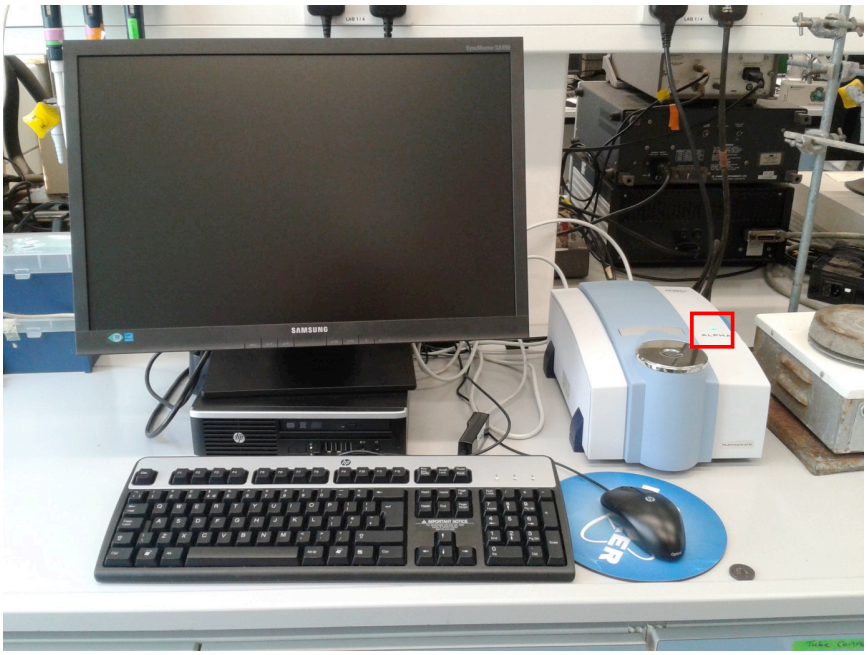


Figure 10: The bench top Alpha spectrometer and accompanying computer. Red box indicates the status light.

8

Figure 11: The position of the prism on the Alpha.

4. Cleaning the prism

Before starting any measurements ensure that the prism is clean and dry. The prism can be cleaned with water followed by ethanol. Ensure all the ethanol has evaporated before continuing. Tissues can be used to wipe the prism but ensure that no tissue fibres are left on the prism surface.

5. Refreshing the background

A background spectrum should be run once a day. If you are unsure when the last background was performed then run a background anyway.

To begin an experiment click the 'REFRESH BACKGROUND' button in the top left-hand side of the main Opus workspace (Fig. 12).

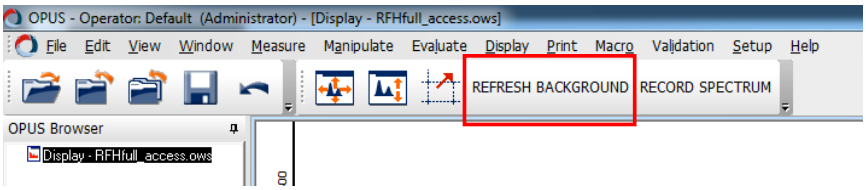


Figure 12: Location of 'REFRESH BACKGROUND' button in main Opus workspace.

A dialog box will open asking whether you want to run a background (Fig. 13). Click 'Continue' to run the background or 'Cancel' to close the window without running a background.

10

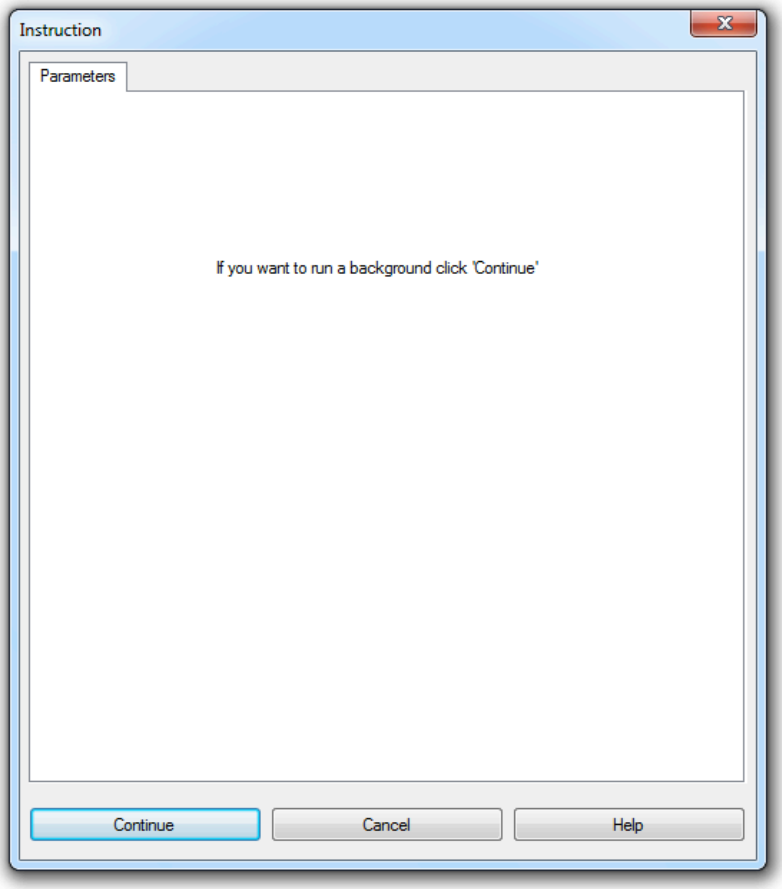


Figure 13: Run background window.

The background takes approx. 2 mins and 4 secs (100 background scans).

6. Recording a sample spectrum

To begin recording sample data click the 'RECORD SPECTRUM' button in the top left-hand side of the Opus main workspace (Fig. 14).

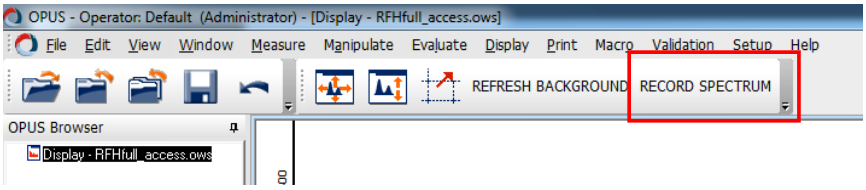


Figure 14: Location of 'RECORD SPECTRUM' button in main Opus workspace.

A dialog box will open asking you to clean the prism (Fig. 15). This should be done at the beginning of the experiment, and after each sample measurement (see section 4).

12

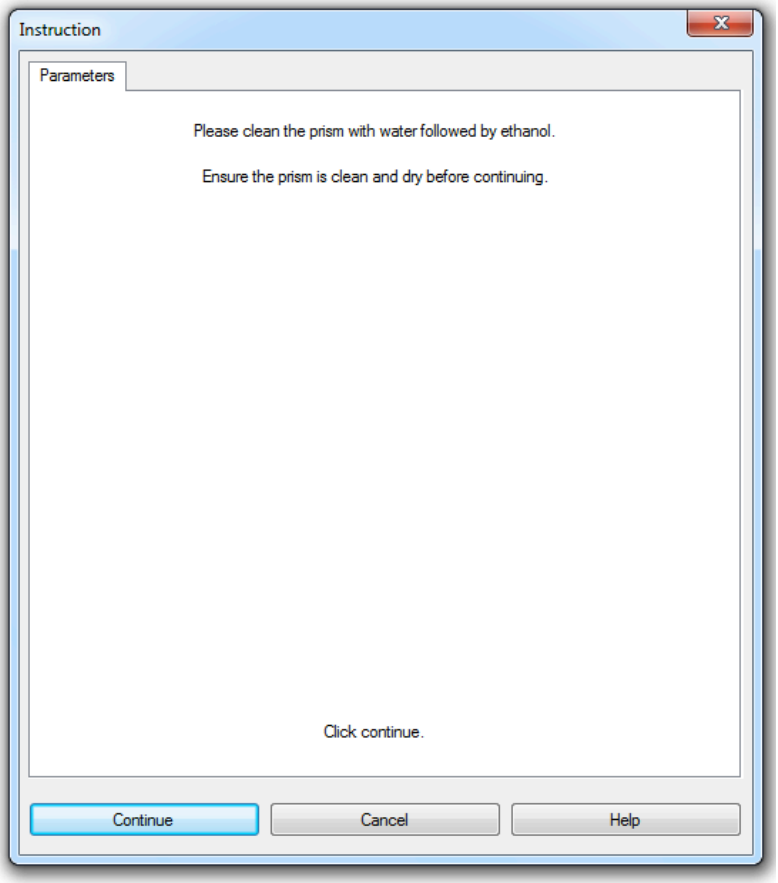


Figure 15: Instruction to clean prism

Once you are happy that the prism is clean and dry, click continue.

You will be then be taken to the next dialog screen (Fig. 16), where you will input a file name. The file name should contain the date that the sample is being recorded followed by an underscore and then a number identifying the sample, which will later be correlated with the patient data. Please do not include any spaces in the file name. There is also a box available to input any additional information that you think should be noted. In most cases, this field can be left blank.

The file name should be in the format:

'mmddyy_IDnumber',

The date should be input in the American format, with month before day, to enable easy sorting of the data into months (it is not the end of the world if the file name is in the incorrect format, but please ensure that the IDnumber is clearly included).

Please keep any additional information short (a few words, no more than 10). Click 'Continue'.

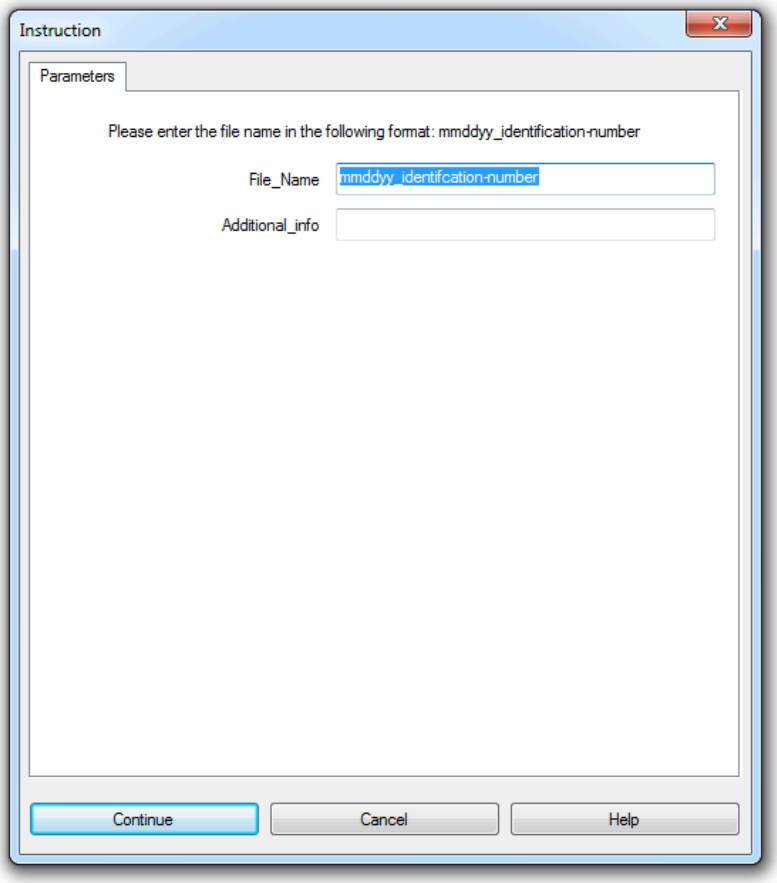


Figure 16: Enter file name and additional information if necessary. Ensure file names are of the format: mmddyy_ID-number.

Another dialogue window opens. Do not load a sample yet. Click 'Continue' to start a live spectrum of the prism surface. When there is no sample on the prism this should be a relatively flat (but noisy) line. If you are happy that the prism is clean then load 50 µL of your sample on to the prism surface, ensuring the entire prism

area is covered. Click the ‘Start Measurement’ button in the bottom left-hand corner of the spectrum window. The sample scan will begin.

If the prism does not look clean click ‘Abort’, clean the prism, and run a new background. When you click ‘Abort’ three error messages will pop up in turn – click ‘Cancel’ on the first and ‘OK’ on the following two.

The status bar at the bottom of the main Opus workspace will be green and a progress bar in the bottom right-hand corner will show percentage completion. Once the scan has finished the status bar will be grey and read ‘No active task’ (See Fig. 8). The spectrum will be unloaded automatically so you will not see any spectrum in the Opus workspace.

The sample scan takes approximately 1 min and 2 secs.

The spectrum is saved automatically. If you want to see the spectrum you can find it in directory:

C:\\OPUS_7.2.139.1294\MEAS\RFH (see section 6 for information regarding loading spectra).

You are now free to continue with further samples. To record another sample simply click the ‘RECORD SPECTRUM’ button and repeat the procedure as before.

7. Loading an Opus Spectrum

This must be done from within OPUS. To load a file you can either click on the shortcut button in the top left-hand side of the main Opus workspace (Fig. 17 top) or go to File – Load File (Fig. 17 bottom).

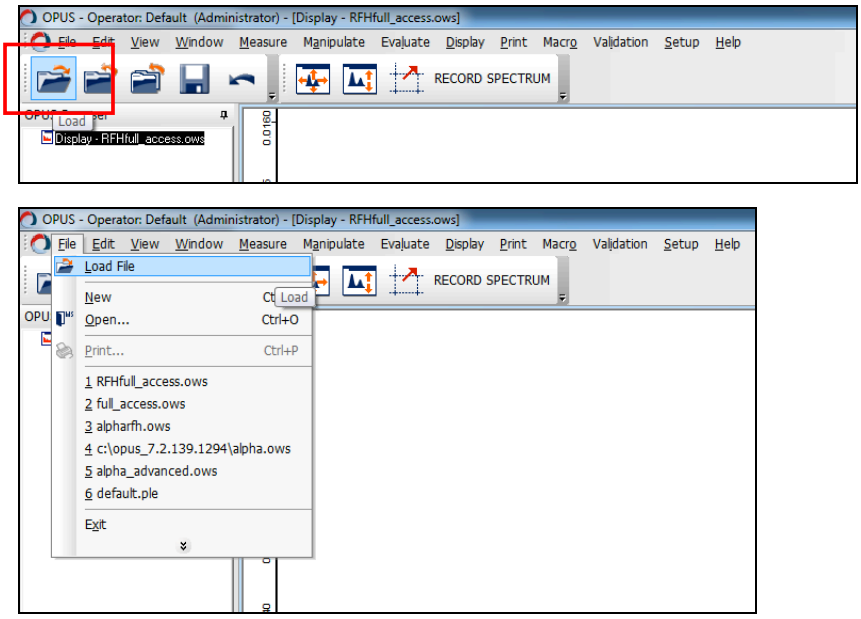


Figure 17: Options for loading a file into Opus, using the shortcut (top) or the file menu (below).

Either option opens up a dialog box and the spectra can be found in (Fig. 18):

C:\\OPUS_7.2.139.1294\MEAS\RFH

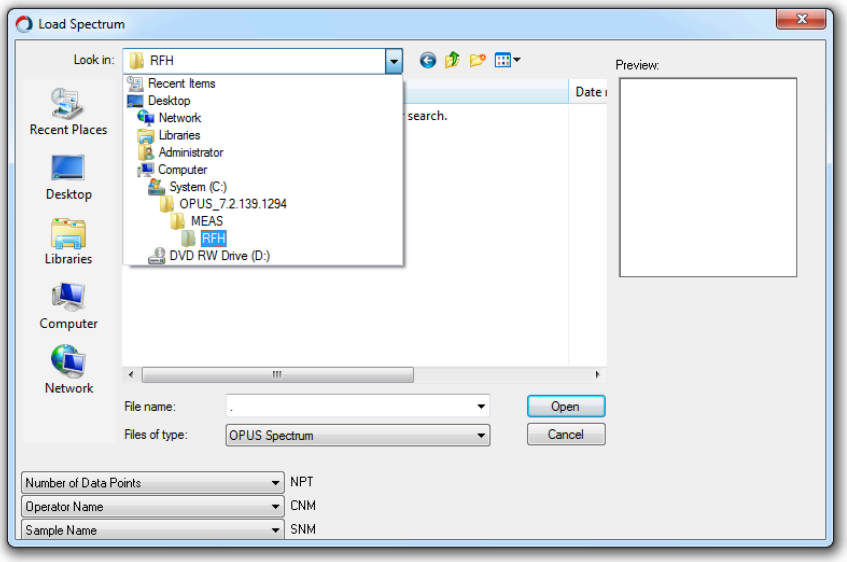


Figure 18: Location of spectra saved after measurement.

Double click on a file to load it into Opus. To unload the file, right click on the file name in the left-hand panel of the main Opus workspace and left click 'Unload File' (Fig. 19). **Please do not save any changes you make to the spectra.**

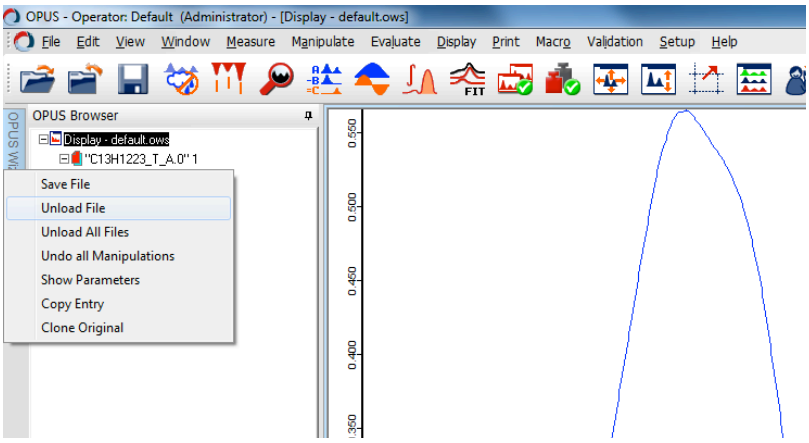


Figure 19: Unloading a file in Opus

8. Troubleshooting

What happens if Opus crashes?

This can happen from time to time. Close Opus by clicking the cross in the top right-hand corner of the window. Optionally restart the computer. If Opus crashes during a measurement and you have to shut down Opus then the spectrum will not be saved. Please repeat the measurement when Opus has restarted.

What happens if I give two files the same name?

Opus does not overwrite files when they are saved with the same name, but instead keeps both files and increases the extension. If you are doing a repeat measurement of the same sample without making any changes then you can keep the file name the same. A note can be made in 'Additional information' about any changes that have been made to the sample. It might also be helpful to note down in a spread sheet any errors made, or email Katie Oliver (contact details in section 8).

Why is the light on the machine not green?

Usually, the light on the machine will be green when everything is OK and it is ready to use. If it is not green then left click the traffic light symbol in Opus (Fig. 5) and check whether any of the components have errors reported. Run any diagnostic tests if expired. If one of the components is reported as faulty or has an error then contact a member of the research team (contact details in section 8).

9. Contact details

In the event of any problems with the machine, computer or software, or if you have any questions please contact a member of the research team.

In the first instance contact:

Katie Oliver

katherine.oliver.11@ucl.ac.uk

Mobile: 07814 606 765

Or

Professor Peter Rich

prr@ucl.ac.uk